Pictogram-Based Eye-Tracking-Driven Robot Control

Usability and Functionality Investigation of Gaze as Control Input and Its Integration into Robotic Interaction Framework

DISSERTATION

for obtaining the doctoral degree

Doctor of Engineering (Dr.-Ing.)

presented by

Anke Fischer-Janzen, M.Sc.

submitted to the School of Science and Technology

of the University of Siegen

Siegen 2026

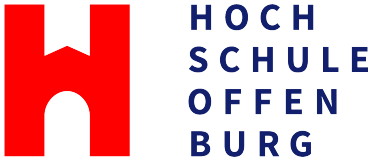
Supervisor and first appraiser
Prof. Dr. Kristof Van Laerhoven
University of Siegen

Second appraiser
Prof. Dr. Thomas M. Wendt
Offenburg University of Applied Sciences

Members of the Doctoral Commission
Prof. Dr. Volker Blanz
University of Siegen
Prof. Dr. Roland Wismüller
University of Siegen

Date of oral examination
17. June 2026





To Bjarne

Acknowledgement

Reaching this point, where my thesis is now complete, has been a long and at times demanding experience, but one that has strengthened me as a researcher and as an individual. Without the people around me, I would not have been able to endure the lows of this dissertation and celebrate the highs. I would like to thank everyone who has walked this path with me, helped me through difficult times, and enriched my work with their own wonderful ideas and thoughts.

This dissertation was written under the supervision of Prof. Dr. Kristof Van Laerhoven, head of the Ubiquitous Computing group at the University of Siegen. I would like to start by thanking him for his motivation in stressful times and close supervision of my work. Thanks to his extensive knowledge, his feedback, and his empathetic and friendly nature, I learned a tremendous deal for my dissertation and for my academic future.

I would like to thank my supervisor and second advisor, Prof. Dr. Thomas M. Wendt, head of the Work-Life Robotics Institute at Offenburg University of Applied Sciences, for his patience during our sometimes hours-long discussions. The freedom he gave me to express myself in my work and develop new skills in my research area has shaped my career. For the excellent and creative brainstorming I'm immensely grateful. His leadership style, characterized by honesty and a tolerance for mistakes, has significantly contributed to my development of self-confidence in my work.

I would like to express my sincere gratitude to both of you, Thomas and Kristof, for giving me the opportunity to pursue my doctorate and for always supporting me throughout these five years.

Thank you to Prof. Dr. Volker Blanz and Prof. Dr. Roland Wismüller from the University of Siegen for agreeing on being my third and fourth examiners as well as taking over the chairing duties within the committee.

I would also like to thank all my colleagues in Offenburg, who have always provided valuable advice and constructive criticism through their expertise. Thank you as well for the fun we had in realizing wonderful projects together that I have thoroughly enjoyed. Special thanks to Alex and Sinan for stepping in for me and supporting me with your administrative work. Furthermore, a special thank you to Niko, Urban, and Philipp for leading the way as pioneers and sharing your experiences with the group. Thank you to Lukas, Steffen, Katrin, and Andi for the numerous discussions and insights that have greatly influenced my work. Additionally, I would like to thank my colleagues in Siegen. I very much appreciated the time I spent in the lab during my visit to Siegen and want to thank you all for the warm welcome.

I would like to thank my friends who put up with me during my doctoral studies and showed understanding for my sometimes weeks-long silence during critical writing phases but also for encouraging conversations when research was exhausting. I would especially like to thank Rebecca for her emotional support over all these years, which goes far beyond the doctoral phase.

Last but not least, I would like to thank my family. I thank my father for sparking my early interest in technology and supporting me throughout my academic career. My brother inspired and encouraged me by making it look so easy to get through an engineering program that I wanted to give it a try myself. And I would particularly like to thank my mother, who has always been a role model for me through her work and commitment and who has supported me in all my career decisions. Thank you to my mother-in-law for her support during the final phase of my doctorate by helping lovingly care for my son. I want to thank my husband Andreas, who endured the long and difficult phases of my doctorate with stoic love and supported me in every way during this process. Thank you for emptying the many dishwashers that I didn't manage to empty myself. Thank you for discussing my papers and listening when I was trying to organize my confused thoughts. Thank you for the emotional support when life was tough, and thank you for our little one, who motivated me to finish this dissertation.

Eidesstattliche Erklärung

Ich erkläre hiermit an Eides statt, dass ich die vorliegende Arbeit ohne unzulässige Hilfe Dritter und ohne Benutzung anderer, nicht angegebener Hilfsmittel angefertigt habe. Die aus anderen Quellen direkt oder indirekt übernommenen Daten und Konzepte sind unter Angabe der Quelle gekennzeichnet. Die Arbeit wurde bisher weder im In- noch im Ausland in gleicher oder ähnlicher Form einer anderen Prüfbehörde vorgelegt. Es wurden keine Dienste eines Promotionsvermittlungsinstituts oder einer ähnlichen Organisation in Anspruch genommen.

Zu den vorgeschlagenen Mitgliedern der Promotionskommission bestehen keine verwandtschaftlichen Beziehungen, keine Verwandschaft ersten Grades, Ehe, Lebenspartnerschaft oder eheähnliche Gemeinschaften.

Anke Fischer-Janzen
Offenburg, den 17. Juni 2026

Abstract

People with severe disabilities strive for a more independence to improve their quality of life. Due to an aging society, the number of people dependent on daily care is expected to increase in the coming years. In response, research in human-robot interaction (HRI) focuses on investigating factors in user experience with robot controls and developing user-oriented robot applications to automate daily tasks. Providing people with severe disabilities with robots that can perform everyday tasks would reduce the workload of caregivers and family members while increasing users' independence.

This work focuses on eye-tracking-driven controls for assistive robotic arms. Using eye-tracking as an input modality offers several advantages: it is noninvasive, requires minimal equipment investment, and allows for quick calibration compared to other input modalities. These benefits facilitate the deployment of such controls. This work identified many approaches to operate a robot with gaze. These approaches differed in the way gaze was employed to control the robot and in the level of effort required to issue commands. A major challenge of eye-tracking-driven controls is achieving three-dimensional control of the robotic arm using two-dimensional eye movements. One solution is to grant the robot more autonomy and enabling it to perform tasks independently. In this case, the user only needs to initiate the desired task. Such an outcome can be achieved through object-oriented gaze or user intent recognition methods. It was found that ambiguous information can result when gaze is tracked on objects that are used in different tasks. Considering how the system is used in everyday life, it is evident that robust task selection is difficult in an unstructured environment with many available tasks. A second challenge in designing eye-tracking controls is instructing the robot to interact correctly with an object. Many approaches use 3D-gaze estimation to determine the object's location relative to the eye-tracker. This requires high accuracy and precise calibration to minimize error.

This doctoral thesis investigated the user experience with eye-tracking-driven controls and feasible control strategy solutions to address the challenges that were identified. The focus rested on creating a scalable control architecture that addresses usability and robustness challenges. This approach overcomes the key limitations identified in object-oriented selection methods. The resulting framework combines gaze-based interaction with semantically structured task representations and architectural safeguards against perceptual uncertainty. The framework allows for free head movement, reducing dependency on precise head tracking and additional sensors, thereby supporting accessible robotic systems.

Keywords: Assistive Robotics, Eye-Tracking, Shared Control, Task-Driven Control.

Zusammenfassung

Menschen mit schweren Behinderungen streben nach mehr Unabhängigkeit im Alltag, wodurch sich ihre Lebensqualität verbessern kann. Voraussichtlich wird die Zahl der Menschen die auf tägliche Pflege angewiesen sind aufgrund einer älter werdenden Gesellschaft steigen. Als Reaktion konzentriert sich die Forschung im Bereich der Mensch-Roboter-Interaktion (HRI) darauf, Faktoren der Benutzererfahrung mit Robotersteuerungen zu untersuchen und benutzerorientierte Roboteranwendungen zu entwickeln, um alltägliche Aufgaben zu automatisieren. Die Bereitstellung von Robotern für die Durchführung alltäglicher Aufgaben würde die Arbeitsbelastung von Pflegekräften und Familienangehörigen verringern und gleichzeitig die Abhängigkeit der Nutzer von anderen Personen reduzieren.

Diese Arbeit konzentriert sich auf Eye-Tracking-basierte Steuerungen für assistive Roboterarme. Im Vergleich zu anderen Eingabemodalitäten bietet die Verwendung von Eye-Tracking mehrere Vorteile: Sie ist nicht-invasiv, ist erschwinglich und ermöglicht eine schnelle Kalibrierung. Diese Arbeit identifizierte mehrere Ansätze, die Blickbewegungen als Eingabebefehl nutzen. Die Ansätze unterschieden sich in der Art und Weise, wie der Blick zur Steuerung des Roboters eingesetzt wird, sowie im Aufwand für die nutzende Person zur Erteilung von Befehlen. Eine Herausforderung bei der Steuerung mittels Eye-Tracking besteht darin, einen Roboterarm mit 2-dimensionalen Augenbewegungen in einem 3-dimensionalen Raum zu steuern. Mit steigender Autonomie können Aufgaben vom Roboter selbstständig ausgeführt werden. Es muss nur die gewünschte Aufgabe von der nutzenden Person initiiert werden. Dies kann durch objektorientierte Blick- oder User-Intent-Recognition-Methoden erreicht werden. Uneindeutige Informationen können generiert werden, wenn der Blick auf Objekte gerichtet wird, die auf unterschiedliche Aufgaben hindeuten können. Die durch den alltäglichen Einsatz entstehenden Einschränkungen machen deutlich, dass eine robuste Aufgabenauswahl in einer unstrukturierten Umgebung mit vielen verfügbaren Aufgaben schwierig ist. Eine zweite Herausforderung besteht darin, dem Roboter beizubringen, korrekt mit einem Objekt zu interagieren. Viele Ansätze verwenden 3D-Gaze-Estimation, um die Position des Objekts relativ zum Eye-Tracker zu bestimmen. Dies erfordert eine hohe Genauigkeit und präzise Kalibrierung des Eye-Tracking-Geräts, um Fehler zu minimieren.

In dieser Doktorarbeit wurden Faktoren der Benutzererfahrung mit Eye-Tracking-basierten Steuerungen untersucht und eine mögliche Lösung basierend auf den gefundenen Herausforderungen entwickelt. Der Schwerpunkt lag auf der Schaffung einer skalierbaren Steuerungsarchitektur, welche die identifizierten Herausforderungen hinsichtlich Benutzerfreundlichkeit und Robustheit löst. Durch die Kombination von blickbasierter Interaktion mit semantisch strukturierten Aufgabenrepräsentationen überwindet der Ansatz wichtige Einschränkungen objektorientierter Auswahlmethoden. Diese Lösungen wurden in Form von Piktogrammen und software-basierten Schutzmaßnahmen gegen Wahrnehmungsunsicherheit umgesetzt. Das entstandene Framework ermöglicht freie Kopfbewegung in der Nutzung, verzichtet auf zusätzliche Sensorik und unterstützt dadurch eine barrierefreie Roboternutzung.

Stichworte: Assistive Robotics, Eye-Tracking, Shared Control, Task-Driven Control.

Contents

1 Introduction **1**
- 1.1 Motivation . . . 1
- 1.2 Problem Formulation . . . 2
- 1.3 Contributions . . . 4
- 1.4 Outline . . . 5

2 Scientific Publications **7**
- 2.1 List of Publications . . . 7
- 2.2 Contribution Statement . . . 9

3 Scoping Review Eye Tracking-based Controls **11**
- 3.1 Introduction . . . 12
  - 3.1.1 Motivation and Research Questions . . . 12
  - 3.1.2 Technical Background . . . 13
- 3.2 Methods . . . 15
  - 3.2.1 Search Strategy and Selection Process . . . 15
  - 3.2.2 Inclusion and Exclusion Criteria . . . 16
  - 3.2.3 Data Charting Process and Data Items . . . 16
  - 3.2.4 Critical Appraisal and Synthesis Methods . . . 18
- 3.3 Results . . . 19
  - 3.3.1 Excluded Publications . . . 20
  - 3.3.2 Related Reviews . . . 20
  - 3.3.3 Included Works . . . 21
- 3.4 Discussion . . . 27
  - 3.4.1 Interpreting Gaze in Robot Control . . . 27
  - 3.4.2 Benefits, Challenges and Limitations of Eye Tracking . . . 28
  - 3.4.3 Future Trends . . . 29
  - 3.4.4 Open Questions . . . 30
  - 3.4.5 Limitations . . . 33
- 3.5 Conclusion . . . 34

4 Requirements of people with disabilities & caregivers **35**
- 4.1 Introduction . . . 36
- 4.2 Methods . . . 37
- 4.3 Results . . . 37
  - 4.3.1 Assisted Living . . . 39
  - 4.3.2 Assisted Work . . . 41
- 4.4 Discussion . . . 43
  - 4.4.1 Comparative Overview . . . 43
  - 4.4.2 Ethical Questions . . . 44

4.5 Conclusion . . . 45

5 WoZ studies on an eye-tracking-driven control 47
5.1 Introduction . . . 48
5.2 Background . . . 49
5.2.1 Gaze as control input in shared controls . . . 49
5.2.2 Characteristics in gaze-based shared controls . . . 50
5.2.3 Tasks in assistive robot manipulation . . . 51
5.2.4 Participation in assistive robot design . . . 52
5.3 Methods . . . 53
5.3.1 Shared control design choices . . . 53
5.3.2 Wizard of Oz design . . . 54
5.3.3 Online survey . . . 55
5.3.4 Hands-on user study . . . 58
5.4 Results . . . 60
5.4.1 Online survey . . . 60
5.4.2 Hands-on study . . . 62
5.5 Discussion . . . 65
5.5.1 Usability . . . 65
5.5.2 Functionality . . . 67
5.5.3 Differences between Online Study and Hands-On User Study . . . 69
5.6 Limitations . . . 70
5.7 Conclusion . . . 70
5.A Overview of Questions asked in the Online Survey . . . 71
5.B Participant statements of hands-on study . . . 73

6 Robust and Scalable Task Selection 75
6.1 Introduction . . . 76
6.2 Related Work . . . 77
6.3 Methods . . . 78
6.3.1 Control strategy . . . 78
6.3.2 Pictogram design process . . . 78
6.3.3 Model selection . . . 79
6.3.4 Data collection, annotation and augmentation . . . 79
6.3.5 Model training and evaluation . . . 80
6.3.6 Initial implementation . . . 81
6.4 Results . . . 81
6.4.1 Task selection . . . 81
6.4.2 Pictogram design evaluation . . . 81
6.4.3 Performance evaluation . . . 83
6.5 Discussion . . . 85
6.5.1 Reflection on key contributions . . . 85
6.6 Gaze control and robot implementation . . . 89
6.7 Limitations . . . 90
6.8 Conclusions . . . 90
6.A Appendix . . . 91

7 Eye-Tracking-Driven Control 93
7.1 Introduction . . . 94

7.2 Related work . . . 95
7.3 Interaction and implementation . . . 97
7.3.1 Human-Robot Interaction process . . . 97
7.3.2 Implementation details . . . 98
7.4 Method exploration . . . 98
7.4.1 Comparison with literature . . . 99
7.4.2 In-the-wild search approach . . . 99
7.4.3 Comparative approach . . . 103
7.5 Framework description . . . 103
7.5.1 Eye-Tracking communication and task selection . . . 103
7.5.2 Human-robot information transfer . . . 104
7.5.3 Shared control interface . . . 105
7.6 Evaluation . . . 106
7.6.1 Test environment . . . 106
7.6.2 Test parameters and data preparation . . . 108
7.6.3 Evaluation Results . . . 108
7.7 Discussion . . . 109
7.7.1 Lessons learned . . . 109
7.7.2 Implementation of robot task accomplishment . . . 111
7.7.3 Limitations and future work . . . 112
7.8 Conclusion . . . 112

8 Vibrotactile Feedback **113**
8.1 Introduction . . . 114
8.2 Methods . . . 115
8.2.1 Vibration perception threshold of the temporal region . . . 115
8.2.2 Measurement of accuracy and precision . . . 115
8.3 Results . . . 117
8.4 Discussion . . . 119
8.5 Conclusion . . . 121

9 Conclusions **123**
9.1 Summary and synthesis of findings . . . 123
9.2 Limitations . . . 126
9.3 Future Work . . . 126

List of Figures **129**

List of Tables **130**

References **133**

Chapter 1

Introduction

1.1 Motivation

As a doctoral student, I had the opportunity to gain insight into the lives of people with disabilities through my collaboration with Lebenshilfe Kinzig- und Elztal e.V. Their interest in the opportunities that robotics offers to improve daily life inspired my work. A key objective for our partners was to free up time for social interactions and communal activities. This idea shaped my goal of designing an intuitive, low-maintenance robotic control system. The question I asked myself throughout this process was what features such a control system should have.

In 2022, estimates indicated that 2.34 million people in Germany lived with a severe physical disability [1]. There are no broken-down figures available worldwide for approximation purposes. Based on previous trend analyses, the incidence of relevant clinical pictures, and an aging society, it can be assumed that the number has risen worldwide since 2022. Concurrently, the shortage of skilled healthcare workers compromises the provision of qualified and attentive care, thereby increasing the workload and stress levels of care staff [2]. Assistive technologies, including assistive robotic arms, can reduce the workload and enhance user independence by daily task execution [3, 4]. A positive user experience increases the likelihood that robotic arms will be accepted and used long-term. This outcome requires adequate functionality and high usability of the control system. In this context, accessibility encompasses more than basic system access. Controls must be adaptable to individual abilities and support the subsequent accomplishment of preferred daily tasks [5, 6]. According to a 2023 interview that I conducted with the robot manufacturers Kinova and Exxomove, only 1,200 robotic arms approved by health insurance companies were in circulation in Germany, Switzerland, and Austria. These robots have been available for several years. The conversations did not address the reasons for the low sales of these models. One possible cause may be limitations in the user experience due to insufficient control variations and input modalities for different types of disabilities.

Unlike other bodily movements, eye movements remain less affected by certain progressive disabilities, such as amyotrophic lateral sclerosis (ALS) [7, 8]. Eye-tracking is a non-invasive control modality with intuitive control strategies that serves as an alternative to joystick controls, head-, sip-and-puff controls, and brain-computer interfaces for people with physical disabilities of the arms and upper body [8, 9, 10, 11]. The precise and fast eye movements, fast calibration process, and low cost of certain eye tracker models advocate for their use as a control input [12]. For these reasons, this approach is also suitable for low- and middle-income countries. Thus, this study focuses on the investigation and design of an eye-tracking-driven control system for robotic arms to assist people with severe physical disabilities in their daily lives. Various approaches to control a robot using gaze are currently being investigated, ranging from simple directional gaze to sophisticated approaches using artificial

intelligence (AI) to detect user intent [12, 13, 14]. The level of autonomy in shared control varies greatly depending on the interpretation method of eye movements. Therefore, shared controls can vary from tele-operated to fully autonomous task execution. Whether a higher level of autonomy correlates with higher user acceptance is debatable, a semi-autonomous control represents a compromise that maintains a sense of agency by sharing control between humans and robots [15, 16]. This can be achieved by using gaze as a selection tool to choose objects or tasks through visual fixation on objects or visual markers. The robot then performs the task autonomously [17, 18, 19, 20]. However, this kind of selection is limited in its ability to scale to a multitude of tasks [21, 22]. Complex tasks, for example pouring a beverage, require more information than just a single selection. For example, the number and location of the objects with which to interact are necessary. Additionally, one object can serve multiple purposes and therefore indicate several tasks, such as a spoon, which can be used for eating, cooking, or setting a table.

In this work, I investigated how shared control influences the user experience, as well as how to overcome challenges found in robust task selection. As a result, I presented EyeRAC, an eye-tracking-driven control for task selection. This open-source framework uses pictograms attached to certain objects with which the user can interact to select tasks. It aims at encouraging other researchers to integrate their solutions for single-task accomplishment into this framework. The level of autonomy may varies between tasks due to the complexity of the trajectory and physical characteristics for each task. While pointing to a certain location sufficed as an input command for some tasks, others required more complex information that had to be processed with deep reinforcement learning approaches. These insights showed that the anticipated gaze-based input is applicable, but solutions for correct selection between multiple tasks and objects must be integrated.

1.2 Problem Formulation

RQ 1. In Germany, the statutory health insurance system provides assistive devices, including robotic arms, to people with disabilities. Before approving a user's request, the users must demonstrate their ability to operate the system appropriately. Systems with easy-to-learn, easy-to-use controls would benefit the process.

But what influences the acceptance and usability of these controls? Challenges in eye-tracking controls can arise from both behavioral and technical sources. Natural gaze behavior is highly reflexive. At the same time, gaze patterns can be used to distinguish between different task intentions based on hand-eye coordination. These partially unpredictable movements must be filtered and analyzed to become predictable to the robotic system. Technological challenges include the limited accuracy of eye-tracking devices due to illumination changes, slippage of glasses, head movements, and calibration errors. Gaze movements, including smooth pursuit and micro-saccades, can result in noise in the tracked gaze data. Although applying filters to reduce noise enhances accuracy, it also leads to information loss. The literature indicates a heterogeneity of research areas in this field that highlight different aspects important for user-oriented designs in eye-tracking-driven controls but overlook others. Due to the multifaceted topic influenced by other modalities, assistive devices, and differences in user preferences and abilities, the literature on user experience with eye-tracking-driven controls leads to heterogeneous results and descriptions, including various characteristics. When I first started researching this topic, there was no summary of existent literature or analysis of influences on the user experience of eye movement-based robot control approaches. In contrast, there was a wealth of literature that focused on individual details of these controls. The following question was asked:

What contributes to the user experience of eye-tracking-driven controls for assistive robotic arms?

RQ 2. The findings of RQ 1 revealed multiple existing controls and indicated different design choices that benefit the usability and functionality of an eye-tracking-driven control. New developments in the field of artificial intelligence (AI) enable the integration of models that allow for new ways to use gaze as an input modality. Regarding shared controls, AI is used to achieve varying levels of autonomy by automating tasks through different learning approaches. Determining the appropriate level of autonomy is an ongoing debate, as it is difficult to attain consensus among participants on what level would maximize user satisfaction. Providing users with a preferred set of functions for daily tasks can lead to satisfaction. Frustration can set in when robots cannot perform tasks or perform them incorrectly. Users must have a sense of agency over the robot, and their cognitive workload must be reduced. This can be achieved through an easy-to-use interface. For example, bounding boxes generated by object detection models can be used as interactive buttons. However, object detection methods face challenges, particularly in unstructured scenes, such as those found in everyday home environments. Object overlap and occlusion, poor lighting, small objects, and a small training set can result in missed detections and mislabeled objects, which may affect the system's functionality. The following question was raised to investigate whether the design meets these requirements.

What usability and functionality challenges arise with object detection-based eye-tracking control?

RQ 3. Recent advancements in user intent recognition research and daily task automation have revealed two major challenges that influence the functionality of control integration when the results of RQ 2 are applied.

First, increasing the number of tasks the robot can perform, also called in this work "scalability", is not always feasible. Most of the reviewed work focuses on accomplishing one to three tasks. One reason is that a system's functionality can be proven with only one task. Other reasons might be that implementing a multitude of tasks is time-consuming and may result in limitations to its reuse by other researchers due to the incompatibility of the libraries or robot models used. Additionally, correctly interpreting gaze movements to extract task intent from random gaze movements is difficult and becomes more challenging with multiple tasks. In the study conducted in relation to RQ 2 it was found that selecting the correct task in the proposed control is challenging due to object geometry, user intent influencing gaze behavior, and technical limitations of the eye tracker itself, which are similar to findings in the literature.

Second, achieving robustness in an environment with multiple tasks is difficult because objects can indicate different tasks, resulting in ambiguous information. This is further complicated by unintentional eye movements. Other studies introduce the necessity of multiple assistive devices to perform different tasks more easily. Combining multiple devices and robustly coordinating their interaction is an ongoing technical challenge. Research Question 3 addresses both challenges.

How can robustness and scalability in task selection be addressed?

RQ 4. The results of the scoping review and the eye-tracking-driven control evaluation indicated that artificial feedback facilitates system usage. Numerous studies have demonstrated the advantages of artificial feedback, which can enhance user experience by offering information and fostering a sense of agency. In this study, auditory feedback was incorporated into the eye-tracking-driven control to improve usability. The review included other studies that used visual feedback through robot motion, screen-based text messages, augmented reality, and light switches to indicate the system's behavior.

Even without additional feedback, the cognitive workload on the visual channel is increased with eye-tracking control. Additionally, freeing the auditory channel for social interactions, such as conversations, improves usability. Auditory cues may limit applicability in certain social situations.

A different approach can be used to maintain the feedback loop. Vibrotactile feedback uses the sense of touch to communicate information to users. For example, discrete signals can indicate successful task selection. The stimulus can be mounted to the eye-tracking glasses to minimize the number of additional devices. The resulting research question is as follows.

What conditions are suitable for applying vibrotactile feedback to the eye-tracking glasses?

1.3 Contributions

The contributions of the works presented in Chapters 3 through 8 summarize the answers to the previously stated research questions. Section 9.1 synthesizes the implications derived from these contributions. The contributions of this work are as follows:

1. Influences on eye-tracking-driven control:
   (a) A systematic overview and comparison of eye-movement-based robotic arm control approaches (Section 3.3).
   (b) Analysis of the reasons for accuracy issues as well as benefits and characteristics that influence acceptance of eye-tracking-driven controls (Section 3.4.2).
   (c) An evaluation of the differences in user satisfaction in works that compare input modalities, limited to those that include eye-tracking controls (Section 3.4.2).
   (d) Identification of research gaps concerning performance measures, subjective perception measurement methods, the inclusion of people with disabilities in the design process, and the application of artificial feedback in the context of eye-tracking-driven robot controls (Section 3.4.4).
   (e) An overview of the needs of people with disabilities regarding eye-tracking controls and robotic arms (Section 3.4.4).
   (f) Derived requirements for assistive robotic systems in assisted living facilities (Section 4.3.1).
   (g) Identification and discussion of the ethical implications of designing and deploying robots in the homes of people with disabilities (Sections 4.4.1 and 4.4.2).
2. Influences on an object detection-based control approach using object-oriented gaze as a selection method:
   (a) Presentation of an initial eye-tracking-driven shared control design that leverages object detection algorithms for daily tasks (Section 5.3.1).
   (b) A functionality and usability evaluation incorporating feedback from potential users, their families, medical professionals, and non-disabled participants with and without prior robot experience (Sections 5.4, 5.5.1 and 5.5.2).
   (c) Refinements and task preferences for the initial shared control design derived from the insights obtained in the two correlating studies (Sections 5.5.1 and 5.5.3).
3. Technical implementation of an eye-tracking-driven control approach:
   (a) An object detection-based eye-tracking control strategy aimed to control robotic arms in daily tasks (Section 6.3.1).

(b) A pilot dataset and an object detection model trained to distinguish eight common daily tasks robustly, systematically evaluated under real-world variability (Section 6.4).

(c) A comparison of feature-matching algorithms tailored to assistive robotic arm use cases used for data transmission (Section 7.4).

(d) An open-source eye-tracking–driven robot control framework based on object detection solving head movement issues by reducing robot control to an eye-in-hand configuration (Section 7.5).

(e) An evaluation of the framework that reveals key challenges and future research directions (Section 7.6).

4. Applicability of artificial feedback methods:

(a) A systematic investigation of the applicability of vibrotactile feedback in the temple region when mounted on eye-tracking glasses (Section 8.3).

(b) Vibrotactile threshold analysis at the temple region (Section 8.3).

(c) Boundary conditions for vibrotactile feedback regarding frequency and amplitude through accuracy and precision evaluation (Section 8.4).

1.4 Outline

The introduction is the first chapter of this doctoral thesis. Section 1.1 explains the motivation behind this work. Section 1.2 presents the problem formulation. Section 1.3 lists the scientific contributions, and Section 1.4 outlines the dissertation.

The second part of the thesis includes the reproduced publications. Chapter 2 is divided into Section 2.1, which introduces the list of publications, and Section 2.2, which presents the contribution statement. Each publication is introduced with a brief summary and details of the authors' contributions. Chapter 3 presents a scoping review summarizing the current state of the art in eye-tracking-based controls for robotic arms. Chapter 4 presents a case study conducted in an assisted living facility and assistive work environment that evaluates the requirements of people with disabilities. Chapter 5 discusses the Wizard of Oz concept, offering suggestions for an initial shared control design. Chapter 6 introduces the pictogram-based control strategy used to communicate user intent to the robot. Chapter 7 describes the resulting eye-tracking-driven control strategy based on previous findings. Chapter 8 is the final publication included in this work and describes the potential application of vibrotactile feedback to explore additional channels for user communication.

Lastly, Chapter 9 is divided into three sections. Section 9.1 summarizes and synthesizes the findings. Section 9.2 discusses limitations, and Section 9.3 presents future work based on the findings of this work.

Chapter 2

Scientific Publications

2.1 List of Publications

This thesis is based on the following publications:

I **A Scoping Review of Gaze and Eye Tracking-Based Control Methods for Assistive Robotic Arms**
Anke Fischer-Janzen, Thomas M. Wendt, and Kristof Van Laerhoven
Frontiers in Robotics and AI, vol. 11, 2024, DOI: 10.3389/frobt.2024.1326670.

II **Requirements of People with Disabilities and Caregivers for Robotics: A Case Study**
Anke Fischer-Janzen, Markus Gapp, Marcus Götten, Katrin-Misel Ponomarjova, Jennifer J. Blöchle, Thomas M. Wendt, Kristof Van Laerhoven, and Thomas Bartscherer
In Preceedings of HCI in Business, Government and Organizations. HCII 2024. Lecture Notes in Computer Science, vol. 14720. Springer, Cham. DOI: /10.1007/978-3-031-61315-9_20.

III **Eye-Tracking-Driven Shared Control for Robotic Arms: Wizard of Oz Studies to Assess Design Choices**
Anke Fischer-Janzen, Thomas M. Wendt, Daniel Görlich, and Kristof Van Laerhoven
Journal of Human-Robot Interaction 15, ACM, vol. 3(56), pp.1-31, 2026, DOI: 10.1145/3796524.

IV **Robust and Scalable Task Selection for Humans and Robots with the Use of Pictograms as Fiducial Markers**
Anke Fischer-Janzen, Thomas M. Wendt, and Kristof Van Laerhoven
IEEE ACCESS, IEEE, vol. 14, pp. 42722-42733, 2026, DOI: 10.1109/ACCESS.2026.3675026.

V **Eye-Tracking-Driven Control in Daily Task Assistance for Assistive Robotic Arms**
Anke Fischer-Janzen, Thomas M. Wendt, and Kristof Van Laerhoven
arXiv:2601.17404, 2026, DOI: 10.48550/arXiv.2601.17404.

VI **Evaluation of precision, accuracy and threshold for the design of vibrotactile feedback in eye tracking applications**
Anke Fischer, Thomas M. Wendt, Lukas Stiglmeier, Philipp Gawron, and Kristof Van Laerhoven
Journal of Sensors and Sensor Systems, vol. 12(1), pp. 103-109, 2023, DOI: 10.5194/jsss-12-103-2023.

All publications are reproduced with permission from their respective publishers.

Publications not included in this thesis:

VII **Shared Control in Assistive Robotics: Characteristics Influencing User Acceptance**
Anke Fischer-Janzen and Thomas M. Wendt
In Proceedings of Sensor and Measurement Science International (SMSI 2026), Nuremberg, Germany, 2026.

VIII **Evaluation of Precision, Accuracy and Threshold for the Design of Vibrotactile Feedback in Eye Tracking Applications**
Anke Fischer, Thomas M. Wendt, Philipp Gawron, Lukas Stiglmeier, and Kristof van Laerhoven
In Proceedings of Sensors and Measuring Systems, 21th ITG/GMA-Symposium, Nuremberg, Germany, 2022, pp. 1-6.

IX **Eye-Tracking Accuracy in Robot Controls: Separating Genuine Issues from Artificial Barriers**
Anke Fischer-Janzen, Julia Roederer, and Thomas M. Wendt
In Proceedings of Sensor and Measurement Science International (SMSI 2026), Nuremberg, Germany, 2026.

X **Accessible Operation Methods of Coffee Machines**
Thomas M. Wendt, Anke Fischer-Janzen, Katrin-Misel Ponomarjova, Kim Zähringer, Andi Schnebel, and Kristof Van Laerhoven.
In Proceedings of the 17th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '24). Association for Computing Machinery, New York, NY, USA, 680–681. 2024, DOI: 10.1145/3652037.3663900

XI **Enhancing Independence through Intelligent Robotics: An AI Driven Assistive Robotics Interface**
Katrin-Misel Ponomarjova, Thomas Wendt, Anke Fischer-Janzen, Sinan Süme, and Bastian Kayser
In Proceedings of UR-AI2023: The Upper-Rhine Artificial Intelligence Symposium: Artificial Intelligence for Time Series, Robotics and Beyond.

XII **WLRI-AD: Assistive Device Dataset for Daily Living Automation**
Katrin-Misel Ponomarjova, Anke Fischer-Janzen, Thomas M. Wendt, and Kristof Van Laerhoven
Personal and Ubiquitous Computing 30(2), 2026, DOI: 10.1007/s00779-026-01854-2

XIII **Innovative and cost-effective Measurement Setup to determine Robot Accuracy**
Markus Gapp, Anke Fischer, and Thomas M. Wendt
In Proceedings of Sensor and Measurement Science International (SMSI 2023), pp. 227-228, 2023, DOI: 10.5162/SMSI2023/D4.4

XIV **Benchmarking Robot and Cobot Performance for an Assembly Application**
Markus Gapp, Anke Fischer-Janzen, David Pfister, Thomas M. Wendt, and Stefan Rupitsch
In Proceedings of Sensor and Measurement Science International (SMSI 2025), pp. 269-270, 2025, DOI: 10.5162/SMSI2025/P19

XV **Robotics for Education: Modular, Age-Appropriate, and Practical Approaches to Inspire the Next Generation and Foster Curiosity and Skills**
Thomas M. Wendt, Jennifer J. Blöchle, Anke Fischer-Janzen, and Andi Schnebel
In Proceedings of Robotics in Education (RiE 2025). Lecture Notes in Networks and Systems, vol 1544. Springer, Cham., 2025, DOI: 10.1007/978-3-031-98762-5_1

XVI **Design of a 3D Printed Force Sensor in the Field of Robotics Utilizing an Embedded Constantan Wire as Sensing Element**
Nikolai Hangst, Jan Muellerleile, Thomas M. Wendt, Lukas Stiglmeier, Philipp Gawron, Urban B. Himmelsbach, and Anke Fischer
In Proceedings of Sensors and Measuring Systems, 21th ITG/GMA-Symposium, Nuremberg, Germany, 2022, pp. 1-7.

XVII **Simulation and Verification of 3D Printed Oscillators for Vibration Energy Harvester**
Philipp Gawron, Thomas M. Wendt, Nikolai Hangst, Lukas Stiglmeier, Urban B. Himmelsbach, and Anke Fischer
In Proceedings of Sensors and Measuring Systems, 21th ITG/GMA-Symposium, Nuremberg, Germany, 2022, pp. 1-4.

2.2 Contribution Statement

	Concept	Implementation	Evaluation	Writing
I	●	●	●	●
II	◕	◑	◕	●
III	●	●	◕	●
IV	●	●	●	●
V	●	●	●	●
VI	●	●	●	●
VII	●	●	●	●
VIII	●	●	●	●
IX	●	●	◕	◕
X	●	◕	◕	●
XI	●	◑	◑	◕
XII	◑	◔	◑	●
XIII	◔	◑	◑	◑
XIV	◔	○	◔	◔
XV	◔	○	○	◔
XVI	○	○	○	◔
XVII	○	○	○	◔

The circles indicate the following:

- ○ Anke Fischer-Janzen did not contribute to this point.
- ◔ Anke Fischer-Janzen was a minor contributor to the work.
- ◑ Anke Fischer-Janzen was a contributor to this work.
- ◕ Anke Fischer-Janzen did a majority of the work.
- ● Anke Fischer-Janzen led and did almost all of the work.

Kristof Van Laerhoven and Thomas M. Wendt were Anke Fischer-Janzen's main supervisors and were involved in all publications, providing guidance and feedback. They also contributed to discussions about the study and framework design. The following publications are reproduced with permission from their respective publishers.

Chapter 3

Publication I: A Scoping Review of Gaze and Eye Tracking-based Control Methods for Assistive Robotic Arms

Summary: RQ 1 was based on an initial literature search that revealed only a few works related to eye-tracking-based controls. The variety of approaches to this topic, which are motivated by human-robot interaction (HRI) and include disciplines such as computer science, robotics, and behavioral science, offer different perspectives on the topic. This variety also makes it difficult to obtain a clear and coherent overview. This underscores the importance of investigating the development process of controls from different viewpoints to create a beneficial system for users. This scoping review was conducted to establish a solid foundation for developing eye-tracking-driven controls that meet user requirements, as well as to identify gaps in the current literature.

The review examined the control of physical robotic arms using eye tracking and was the first of its kind in this field. Certain robot applications, such as those in surgical or industrial use, were excluded in order to present comparable data. The main focus was on accomplishing everyday tasks to benefit people with disabilities. Robot types other than robotic arms, such as social robots, were excluded due to heterogeneity in application and mode of interaction.

Forty-five works met the inclusion criteria. These findings influenced the development of the eye-tracking control described in Chapter 7. Specifically, the implementation methods of shared control approaches, performance measures, and perception characteristics influenced the design. A lack of participation opportunities for people with disabilities was identified. The distinct expertise and perspectives of people with disabilities, which differ from those of non-disabled users, highlighted the importance of examining the user experience of such controls with stakeholder groups. These observations motivated the concurrent usability investigation presented in Chapter 5. A case study was conducted in Chapter 4 to gain deeper insight into user needs regarding robotics in everyday life. The findings suggested usability benefits from applying artificial feedback in human-robot interaction. An evaluation of vibrotactile feedback is presented in Chapter 8.

Contribution: Anke Fischer-Janzen was the primary author of this publication. She was responsible for its conception, implementation, evaluation, and preparation. Thomas M. Wendt and Kristof Van Laerhoven contributed to the project by providing supervision, methodological discussions, and reviewing the manuscript.

A Scoping Review of Gaze and Eye Tracking-based Control Methods for Assistive Robotic Arms

Anke Fischer-Janzen
Offenburg University of Applied Sciences
Anke.Fischer-Janzen@hs-offenburg.de

Thomas M. Wendt
Offenburg University of Applied Sciences
Thomas.Wendt@hs-offenburg.de

Kristof Van Laerhoven
University of Siegen
kvl@eit.uni-siegen.de

Abstract

Background: Assistive Robotic Arms are designed to assist physically disabled people with daily activities. Existing joysticks and head controls are not applicable for severely disabled people such as people with Locked-in Syndrome. Therefore, eye tracking control is part of ongoing research. The related literature spans many disciplines, creating a heterogeneous field that makes it difficult to gain an overview.

Objectives: This work focuses on ARAs that are controlled by gaze and eye movements. By answering the research questions, this paper provides details on the design of the systems, a comparison of input modalities, methods for measuring the performance of these controls, and an outlook on research areas that gained interest in recent years.

Methods: This review was conducted as outlined in the PRISMA 2020 Statement. After identifying a wide range of approaches in use the authors decided to use the PRISMA-ScR extension for a scoping review to present the results. The identification process was carried out by screening three databases. After the screening process, a snowball search was conducted.

Results: 39 articles and 6 reviews were included in this article. Characteristics related to the system and study design were extracted and presented divided into three groups based on the use of eye tracking.

Conclusion: This paper aims to provide an overview for researchers new to the field by offering insight into eye tracking based robot controllers. We have identified open questions that need to be answered in order to provide people with severe motor function loss with systems that are highly usable and accessible.

Keywords: robot, assistive robotics, input modalities, eye tracking, assisted living, EOG, hybrid BCI, human robot interaction (HRI)

3.1 Introduction

3.1.1 Motivation and Research Questions

Assistive robotics is a broad field that describes the use of robots to assist the elderly or people with physical and cognitive disabilities. The field describes several types of robotic applications. For example, social, service and surgical robots, walking aids, exoskeletons, power wheelchairs, therapy robots and assistive robotic arms. The term Assistive Robotic Arm (ARA) is only one of several keywords used in the literature. The rationale for this work is based on the versatile naming found in various publications. Robotic arms used to assist people in everyday life have been called, among others, wheelchair-mounted robotic arm, assistive robotic manipulator and assistive robot, which complicates the retrieval of related work. In this paper, we will use the term Assistive Robotic Arm.

Controlling a robot with eye movements could be an appropriate solution for people with severe physical impairments of arms, head movement, and speech. In Germany, for example, 7.8 million people are severely disabled. 11% have disabilities of the arms and legs, 10% of the spine and upper body, and 9% have a cerebral disorder [1]. This results in approximately 2.3 million people who

could benefit from an eye tracking control. Locked-in patients in particular could benefit from an ARA. Locked-in Syndrome can result from brainstem lesions such as stroke, traumatic brain injury, and tumors, from brainstem infections or degenerative diseases. Depending on the individual case rehabilitation therapy can restore body functions [23]. Amyotrophic Lateral Sclerosis (ALS) is a progressive disease that immobilizes the patient over time. In the Locked-In state and in the late stages of ALS, there is a high probability that the motor function of the eye is still intact, showing possibilities for the use of eye tracking [7].

Currently, robots are used primarily in industrial applications to automate work. Adapting such robots as assistive systems requires interdisciplinary knowledge due to the variety of daily tasks and environments. Such tasks can be grouped into Activities of Daily Living (ADL) [24]. They involve routines such as cleaning, eating, and personal hygiene, to ensure a good quality of life. Today, ARA can be controlled by joysticks. In these cases, the 3-dimensional moving robot is controlled solely by the user. Shared control, where Artificial Intelligence (AI) is used to reduce the cognitive load of the user in complex tasks, is being explored [25, 26]. However, concerns about Human Robot Interaction (HRI) and safety are raised due to the close proximity to the user [26]. A concern in assistive robotics is to ensure the well-being of the user. Among the challenges of realizing the various applications and ensuring the safety of the user, the usability and accessibility of ARAs must be addressed. If the system is not easily accessible and adapted to individual needs, people tend not to use it. A related challenge is the Midas touch problem, which describes the misinterpretation of gaze that triggers robot commands. This leads to user frustration [27, 28]. This challenge is compounded by the need to control the robot in a 3-dimensional space using 2-dimensional eye movements. For these reasons, this work will address the research questions in Table 3.1.

The main contributions of this work are based on the answers to these questions.

1. An overview of publications in the heterogeneous field of eye movement based robot control.
2. A detailed specification of the technology used (link) and the studies conducted (link)
3. Future trends and open questions in eye-movement based robot control and a comparison of approaches

Table 3.1: Research Questions and motivation

No.	Research Question
RQ1	What approaches have been explored in the field of gaze-controlled robotic arms to assist people with (severe) upper limb impairments?
RQ2	How can eye gaze be interpreted to control a robot?
RQ3	What are the developments, limitations and challenges that can be found in the field of gaze-controlled assistive robotic arms?

3.1.2 Technical Background

Eye movements are used in a variety of assistive technology applications. Popular are eye typing interfaces that provide the user with the ability to speak and eye mouse that allows the person to use a computer [29]. Other applications include the control of electric wheelchairs that provide mobility to the user[30]. Social robots use it to extract facial features and interpret the feelings and needs of the user [2].

Fixation and saccades are the most common eye movements. Fixation describes the gaze resting on a particular point. The duration of fixation depends on the object or location being fixated and ranges

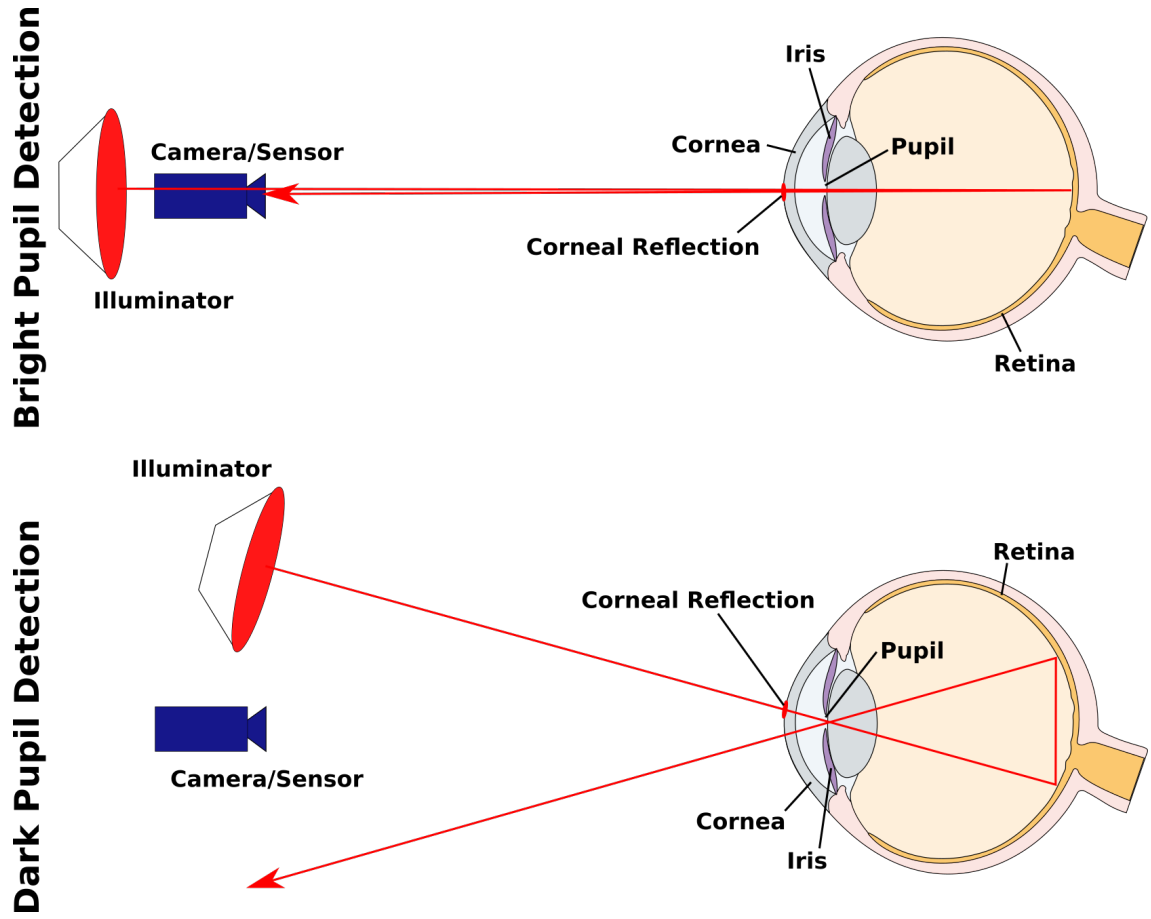


Figure 3.1: A comparison between bright pupil and dark pupil detection. The corneal reflection used for determining the Pupil Corneal Reflection Point is shown on the cornea. Adapted from Tobii Connect [32], the anatomy is presented in a simplified form.

from tens of milliseconds or seconds. Fixations can be measured as dwell time, which is the amount of time a user fixates on an object. This is often used as a parameter to interpret the user's intent or to give a command to the robot. Saccades are rapid movements of the eye. They last only a few milliseconds and can be as fast as $500°/s$ [31]. These numbers vary in the literature literature. This may be due to the fact that everyone is anatomically and behaviorally unique, including how they move the eyes.

As the environment and the person move or the depth of field changes, the eye responds with vergence movements and smooth pursuit. Vergence movements describe the movement of the eye during reading. The axis of the eyes moves from parallel (distance vision) to an intersection on the page (near vision). This prevents double vision. Smooth pursuit describes the eye movement while looking at an object and moving the head or the object itself. This eye movement is not voluntarily. The slow speed of the eye with less than $30°/s$ in smooth pursuit distinguishes it from saccades [31]. Most of the systems evaluated use fixation on objects or directions, but in head-mounted eye tracking devices smooth pursuit could have an impact on the function of the system.

There are several techniques to measure these movements. The techniques can be divided into Infrared Oculography (IOG), Video Oculography (VOG), Electrooculography (EOG), and electromagnetic coils [33]. All techniques are used to measure vertical and horizontal eye movements. Depending on the system, additional movements such as eyebrow movements, blinks, and pupil dilation can be recorded.

IOG uses infrared sensors to detect IR light. Unlike VOG, which records visible light, the invisible light is less distracting to the user, especially in darker environments. Most eye trackers require additional illuminators to create the following effect. When using Pupil Corneal Reflection Point effect, illumination creates a reflection on the cornea. By tracking the vector between the center of the pupil and this reflection eye movement can be determined [33]. As it will be presented, most eye trackers are based on infrared light because it provides more contrast of the pupil. This is based on the dark pupil detection method, where by illuminating the eye with IR light, the pupil appears darker than the iris [7]. Another way to achieve this effect is to place the illuminator at a different angle to the camera (see Figure 3.1). With this method, the light reflected from the retina is blocked by the iris, resulting in a darker pupil. In contrast, for bright pupil detection the illuminator is placed close to the optical axis of the camera, creating what is known in photography as the red-eye effect. The light reflected from the retina is captured by the camera [32]. Depending on the algorithm, the appearance of the pupil contrast can lead to better measurement results. With IOG and VOG it is possible to measure the pupil size in addition to the horizontal and vertical movements of the eye, as is done in behavioral studies [32].

EOG measures corneo-retinal standing potentials by detecting the electric field and is used to measure horizontal, vertical eye movements and eyebrow movements. This method is based on placing multiple electrodes on the user's face in specific regions. Changes in the electric field are generated by the eye movements [33]. These systems can be used with a hybrid Brain Computer Interface (hBCI). A Brain Computer Interface (BCI) uses Electroencephalography (EEG) to measure neural signals. EEG is a mostly non-invasive neuroimaging technique used to measure and record the electrical activity generated by the brain through electrodes placed on the scalp. These signals can be used to interpret the user's intention and to control external devices such as speech computers and robots [34, 35].

hBCIs can use additional inputs such as EOG to improve the quality of the measurement, as will be presented in the results. For example in Huang et al. [36] Steady-State visually Evoked Potential (SSVEP) is used to control the system. SSVEP is an electrical signal evoked by the brain's response to a visual flickering stimulus that has a constant frequency. From this pattern, the brain generates generative oscillatory electrical activity that can be measured by EEG.

Since no related publications using electromagnetic coils were found in this work, we refer to Holmqvist and Andersson [31], Klaib et al. [33], and Edughele et al. [7] for further information. These techniques are used to classify the retrieved publications as shown in Table 3.4. There will be occasions when a system is labeled as VOG/ IOG. No detailed constraints could be found in literature for the eye tracker technique used. This is possible with both infrared and visible light.

3.2 Methods

This review was conducted in accordance with the PRISMA 2020 Statement. Due to the heterogeneity of the literature the PRISMA-ScR extension [37] was used to present the results. In terms of the research questions, this review focuses on eye tracking based control for robots. In the following the databases, search terms, inclusion and exclusion criteria are listed below.

3.2.1 Search Strategy and Selection Process

Three databases listed in Table 3.2 were searched to identify publications of interest. During the planning phase, non-standardized Internet searches were conducted as a preliminary evaluation to gather further information and identify search terms, including commercial systems and patents. After the systematic identification and screening process, a snowball search was conducted to identify relevant publications from other databases such as Scopus. Publications in English and German were included in the review process.

The search terms used for identification were permutations of the words "eye tracking", "robot", "eye", "shared control", "gaze" and "assistive". Combinations of two search terms such as "eye" AND "robot" were neglected due to the number of publications found from other research topics outside the scope. As can be seen in Figure 3.2, the vast majority of reports were manually excluded due to ambiguous terms. Searching for "assistive robotic arms" (ARA) or "wheelchair mounted arms" (WMRA), which would imply an aid for physically disabled people, would have excluded most of the relevant papers. The general use of these terms is not common and varies between disciplines. Similar results can be found for eye and gaze used as synonyms, leading to ambiguous results for robotics, such as "eye-in-hand" describing the use of a camera mounted on the robot's end effector to improve automated grasping.

Table 3.2: Databases used in the identification phase.

Data base	Available at	Last accessed
IEEE	https://ieeexplore.ieee.org/Xplore/home.jsp	07.07.2022
ACM	https://dl.acm.org/	01.02.2023
MDPI	https://www.mdpi.com/	01.02.2023

3.2.2 Inclusion and Exclusion Criteria

Table 3.3 shows the inclusion and exclusion criteria. Publications were included in the evaluation if the robot was controlled by eye movements. Accordingly, robot type and eye tracking method were included as evaluation items in Table 3.4. Any system used in a non-assistive task was excluded. This decision led to the inclusion of industrial applications with shared workspaces, as the system could also be used by people with motor impairments in their working life. Other excluded applications can be summarized as surgical assistance systems and telepresence robots in hazardous environments. In these cases the robot is not specifically designed to assist people with physical disabilities. Rehabilitation robots are used to mobilize the person. In such tasks, the robot's movements are adapted to meet the patient's anatomy. Visual feedback can be helpful to improve the rehabilitation process. However, as it will be shown the application of eye tracking varies greatly and many systems are exoskeletons. An exoskeleton is not controlled by gaze alone. To ensure safe use, torque, force, and pressure sensors are used to prevent unhealthy forces on the body. Rehabilitation robots are usually limited to a certain movement set, which is different from the wide range needed in everyday life. Therefore, the architecture of a rehabilitation system is different from an ARA, making a valid comparison impossible. The first research question depends on several variables: the space in which the robot can move, the level of automation and whether the systems presented have been tested by users (able-bodied or disabled). If no information was provided, this did not lead to exclusion due to the small number of publications included.

3.2.3 Data Charting Process and Data Items

The data chart presented in Table 4 and online was constructed iteratively. Variables were selected with respect to author information, hardware and software setup, and empirical study criteria. The rationale for the chosen parameters was to find parameters that would allow the reader to gain insight into the design and development of eye tracking based ARA and, if already developing such a system, to find similar approaches. While reading the included publications the importance of the selected parameters for the original authors was estimated and the table was adapted. This resulted in two tables with 17 (link) and 12 (link) elements respectively. Since these tables were too large to fit in

Table 3.3: Inclusion and exclusion criteria

No.	Inclusion Criterion
1	A robotic arm is controlled by eye tracking, regardless of the type of robot, the eye tracking method, or the area of use.
2	Eye tracking is used as a control input for a robotic arm among other control inputs.
No.	**Exclusion Criterion**
1	The type of publication was a bachelor's, master's, or doctoral thesis, a patent, or a commercial system.
2	When eye tracking was used to control something else than a robotic arm, or the control of the robotic arm was only implied.
3	Eye tracking was used as a monitoring device only.
4	Publications in which the robot was used as an exoskeleton or as a therapeutic robot.

this paper, it was decided to extract the most important ones, presented in Table 3.4, to facilitate the reading of the paper. For the author information four items were selected (names, year, title, and DOI). The system information was divided into six items (technology overview, eye movement detection device, eye tracking technique and sensors, wearable or remote eye tracking device, robot manipulation space dimensions, and type of robot). The software was described in three parameters (algorithms and models used in the approach, programming environment, type of eye movement used to control the robot). Finally, the empirical parameters were specified in four parameters (task description, number of participants, number of repetitions, empirical test performed) and additional three parameters describing the specified measurements and parameters divided into task related, computational and empirical parameters. This list was prepared by the first author and discussed with all authors.

Table 3.4: List of articles included. Eye tracking devices were divided intoVOG, IOG and EOG. Robotic arms were distinguished between their use in assisted living (ARA), industrial (R), educational (EDR) or collaborative robots (CR). Eye movements are described as fixations (F), saccades (S), blinks (B) and winks (W) and SSVEP (E). The last three columns describe the number of able-bodied (AP) and disabled participants (DP), and trials (T) in the user study. More information can be found in the tables provided in Fischer-Janzen et al. [38].

Author, Year	Technology		Motion	Eye Movem.	Robot Space	AP	DP	T
			Telemanipulation					
[39]	VOG	Head-worn	R	Point	F	3D	-	-	-
[40]	IOG	Remote	n.s.	Point	F	3D	-	-	-
[26]	IOG	Head-worn	R	Task	n.s.	n.s.	-	-	-
[41]	IOG	Head-worn	R	Task	F	3D	0	6	20
[42]	EOG	Head-worn	R	Point	S	2D	3	0	n.s.
[43]	IOG/VOG	Remote	PR	Grasp	F	3D	2	0	28/31
[44]	IOG	Remote	ARA	Grasp	F	3D	5	0	5
[45]	IOG	Remote	R	Task	F	3D	8	0	30
[46]	VOG	Remote	n.s.	Task	F	2D	20	0	9
[36]	EOG	Head-worn	ARA	Task	B+E	3D	5	0	3
[47]	IOG	Remote	EDR	Task	F	3D	9/6	9/6	2/2

[48]	IOG	Head-worn	CR	Task	F	3D	10	1	33
[28]	IOG	Remote	R	Point	F	3D	7	0	3
[49]	IOG	Remote	CR	Task	F+S	3D	-	-	-
[50]	IOG	Remote	CR	Task	F+S	3D	-	-	-
[51]	IOG	Remote	R	Task	F	3D	10	0	15
[52]	IOG/VOG	Remote	CR	Task	F	3D	-	-	-
[53]	VOG	Remote	EDR	Task	F	n.s.	7	6	1
				Directional Gaze					
[54]	IOG/VOG	Head-worn	R	Task	F	3D	-	-	-
[55]	EOG	Head-worn	R	Point	F	2D	6	0	n.s.
[56]	EOG	Head-worn	R	Task	B	3D	-	-	-
[57]	EOG	Head-worn	R	Point	S	2D	4	0	n.s.
[58]	IOG/VOG	Remote	n.s.	Grasp	F+P	n.s.	8	0	n.s.
[27]	IOG/VOG	Head-worn	CR	Task	F+B	3D	10	1	1
[59]	IOG	Remote	CR	Task	F+W	3D	8	0	5
[60]	IOG	Remote	EDR	Point	F	2D	8	0	30
[61]	EOG	Head-worn	n.s.	Point	F	3D	30	n.s.	20
[62]	EOG	Head-Worn	n.s.	Point	F	2D	3	0	20
[63]	EOG	Head-worn	n.s.	Point	F+B	2D	3	0	n.s.
				Object-oriented Gaze					
[64]	IOG	Head-worn	ARA	Task	F	3D	0	9	var.
[65]	IOG/VOG	Head-worn	ARA	Grasp	n.s.	3D	26	0	1
[17]	IOG	Remote	EDR	Grasp	F+B	3D	7	0	n.s.
[18]	IOG	Head-worn	ARA	Grasp	F	3D	-	-	-
[66]	IOG	Head-worn	ARA	Grasp	F	3D	n.s.	0	4
[67]	EOG, IOG	Head-worn	ARA	Grasp	F	3D	10	0	20
[19]	VOG	Remote	ARA	Grasp	F	3D	-	-	-
[68]	IOG	Head-worn	CR	Point	F+B	3D	3	0	10
[20]	IOG	Head-worn	PR	Point	F	3D	5	0	5
[69]	IOG/VOG	Head-worn	CR	Grasp	F	3D	-	-	-

3.2.4 Critical Appraisal and Synthesis Methods

The interpretation of the results is limited to the statements of the cited literature. The list may not be exhaustive for research areas beyond eye tracking techniques in the context of robotics based on the search terms. As indicated by several authors, telemanipulation of robots includes various control input devices. Related research on input modalities such as joysticks may provide different perspectives to the discussion presented here. A risk of bias may exist for some of the evaluation items. Multimodal control and shared control exist for other areas of research. For example, gaze-based control is increasingly used to control electric wheelchairs and computer interfaces. An in-depth look at algorithmic design proposals is beyond the scope of this paper and not feasible due to the heterogeneity of the included work. However, the question arises to what extent the results are transferable. In order to reduce the risk of bias, reviews on the aforementioned topics were reviewed, interpreted, and included in the discussion.

The retrieved data will be evaluated in terms of challenges, benefits, limitations, and identified research interests. The challenges, benefits, and limitations are extracted and generalized in the case of duplicates. Sources will be provided for each summarized item. The research interests are extracted

by sorting the publications according to their aim and contribution. Three timelines were created for three categories: telemanipulation, directional gaze, object-oriented gaze. Keywords were evaluated to indicate category membership.

The presentation of articles varied greatly depending on whether the researcher was working in computer science, robotics, or behavioral science. Therefore, research interests were extracted through a summary of keywords resulting from these parameters to reduce the influence of personal and field bias. Author defined keywords were collected and ranked according to their frequency of occurrence.

3.3 Results

The adapted PRISMA 2020 state flow is shown in Figure 3.2. 3227 publications were identified and screened for duplicates. This was followed by title and abstract screening. Duplicates were found by applying a self-written Matlab script. Duplicates found later were added to the number of "duplicate records removed" accordingly. After screening the abstracts, 62 publications remained. These full texts were checked against the exclusion criteria. A snowball search based on the cited authors and an Internet search of the lead authors in each publication yielded 16 additional publications of interest. A total of 39 articles and 6 reviews were included in this work.

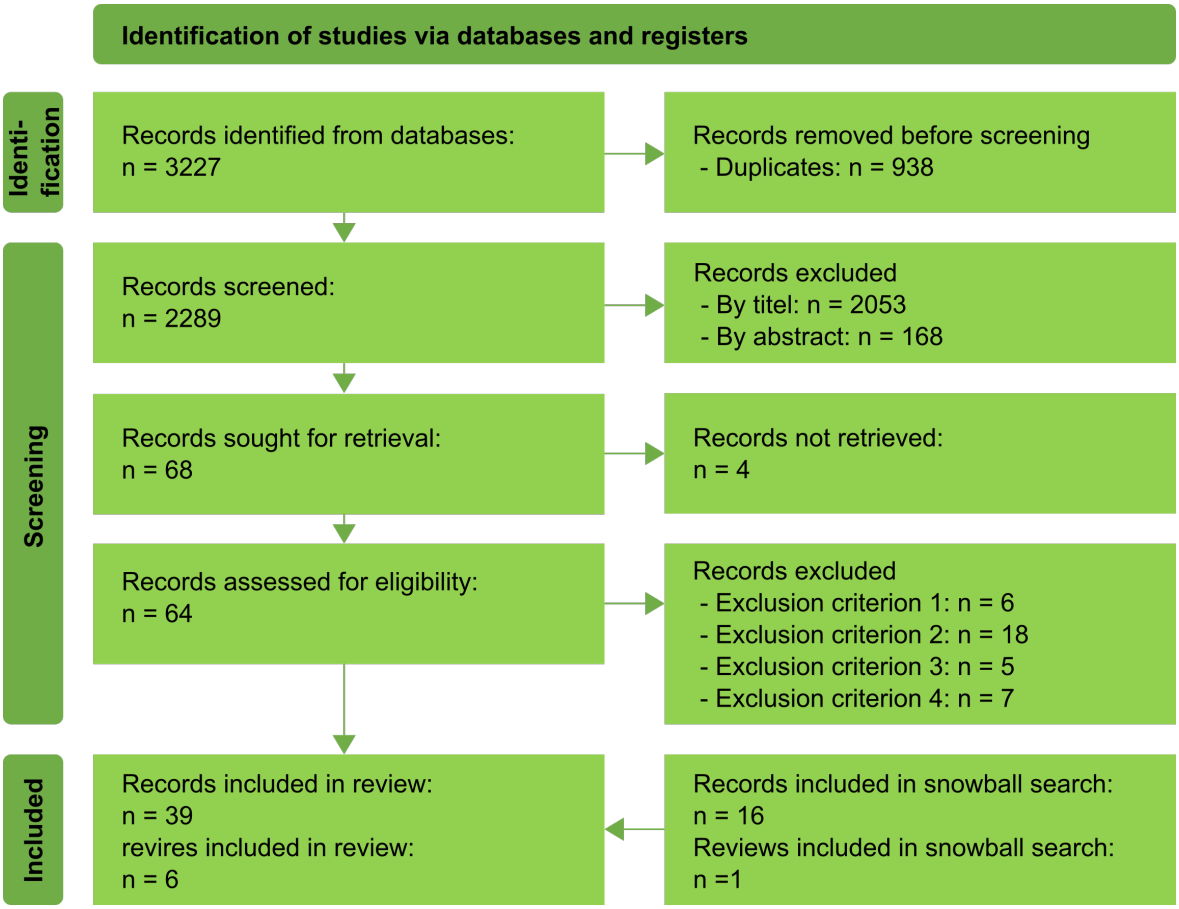


Figure 3.2: The state flow shows the identification and screening process according to PRISMA 2020.

3.3.1 Excluded Publications

Due to the rapid progress in research topics related to robotics and eye tracking, this section provides a brief overview. Especially telepresence robots, social robots, and the use of hBCI and BCI were found to be of growing research interest. Rehabilitation and surgical assistance were mentioned as possible applications. The 36 excluded publications were used to identify these topics.

Six publications were excluded due to their publication type. Three of these were dissertations. Alsharif's Ph.D. thesis can be partially represented in her publication, which is referred to later in the included articles [70]. Shahzad et al. [71] presented a master's thesis on controlling articulated robot arms using eye tracking. Nunez-Varela [72] published a Ph.D. thesis in 2012, focusing on task-driven object manipulation by gaze locations on the object (intention read from gaze). The MyEccPupil (HomeBrace GmbH) is a commercially available eye tracking controller for a wheelchair-mounted robotic arm. Although other eye tracking based systems were searched worldwide, no other systems could be found. Two related patents were found [73, 74].

Telepresence robots [75, 76], smart or electric wheelchairs [77, 30, 7, 78, 79, 80, 81, 82], social or humanoid robots [2, 83, 84, 85, 86, 75, 87] are controlled by gaze. Such programs greatly facilitate the daily life of people with upper limb impairments. The movement of other devices such as cameras by gaze can help in the control of surgical robots for laparoscopy [9, 88]. Similar to direction corrections as in surgical context gaze can be used to interpret human intentions. Human intend detection were used to interpret the desired action the robot should accomplish to enhance the performance of a shared control [89]. A human reflex also includes the dilatation of the pupil in certain events and can be used for shared control improvement [90, 91, 92, 25].

Eye tracking is a versatile technology and is used, among other things, to detect whether the participant's gaze is fixated on the monitor presenting the robot control interface [93]. Projects such as ASPICE focus on a selection of input modalities to match the needs of the participants in order to assist patients with different tasks in the form of a hBCI [94]. A robotic dog (AIBO) was used to test the modalities.

A variety of robotic applications can be found in rehabilitation. In cases of stroke rehabilitation, the robots are used to move the extremities through gaze implication, which is detected by EEG, EOG and VOG [95, 96, 97, 98, 99]. Currently, it has been shown to restore hand function, but the "fluent, reliable and safe operation of a semi-autonomous whole-arm exoskeleton restoring ADLs" have to be demonstrated [95]. Approaches for home physical therapy can be realized by exoskeletons [96, 99, 100] or robotic gloves [97, 98, 101]. Assisting with drinking tasks using an exoskeleton is mentioned in Crea et al. [95]. Other tasks include food preparation using virtual reality and an exoskeleton [102] or drawing on a screen without a robotic system [103].

3.3.2 Related Reviews

This section provides a brief overview of reviews in related research areas. A review of input modalities has been conducted by Clark et al. [104], who present eye tracking, computer vision and EEG approaches, emotion recognition, gestures, and lie detection. Advantages and challenges are presented along with an extensive literature review. It is also stated that non-verbal and non-touch based approaches are important for the future development of intuitive and natural feeling robot control. Esposito et al. [105] reviewed biosignal-based Human Machine Interaction (HMI). These were biopotential biosignals such as EEG, muscle-mechanical motion, body motion, and hybrid approaches. This review focused on the use of EOG, among others. Ramkumar et al. [106] presented an overview of the review classification of EOG-based Human Computer Interfaces (HCI) considering data from 2015 to 2020. In addition, Schäfer et al. [107] compared five hands-free input modalities that are important in robotic arm control research. Dünser et al. [108] presented a similar approach and compared four input modalities. The results of both works are discussed in Section 3.4.3. In addition, a comparison

of BCI and eye tracking in eye typing studies using a spelling program with people with severe motor impairments [109].

Robotics in healthcare is a promising approach addressing the shortage of skilled workers. The goal is to support patients in physical therapy and everyday life as well as caregivers in their daily tasks. Human recognition, emotion and speech recognition are used for the realization. Kyrarini et al. [2] present robots in different scenarios, such as care robots like Pepper, hospital robots that help with logistics, physical therapy robots like exoskeletons and walking aids and finally assistive robots like FRIEND, Jaco 2, and Baxter.

Finally, a review of wearable interaction was found that defined head and eye movements used to control wearables [110]. The results showed that there is limited research on the accessibility of such systems for people with motor impairments. The main findings of this review revealed four suggestions for future research. 1. exploring a variety of wearables, as the current focus is on head wearables, 2. multimodal input modalities and input modalities that maximize motor abilities, 3. more user studies, and 4. IoT.

3.3.3 Included Works

The literature found is the basis for answering research question RQ1: "What approaches have been explored in the field of gaze-controlled robotic arms to assist people with (severe) upper limb impairments?". Table 3.4 lists all articles that met the inclusion criteria. It is divided into three sections called telemanipulation, directional gaze, and object-oriented gaze. This separation was chosen because the interaction between the participant and the system can vary. Telemanipulation describes the use of a display that allows visual feedback and the presentation of buttons to manipulate the robot. Directional gaze approaches are based on looking in a direction to move the robot. Switching between robot manipulation dimensions (e.g. x-y axis to x-z axis) or grasping mode is achieved by including blinking or using additional input modalities. Object-oriented gaze entails the need to either determine gaze in 3D-space, implement object detection algorithms, and adapt trajectory planning to automate the task. For all three categories, we distinguish between 1) assistive systems in everyday life and 2) miscellaneous applications, which include industrial applications, comparative work, and tasks that are not performed in everyday life.

In Table 3.4, the Technology column is divided into the type of eye tracking device used, the location of the device, and the type of robotic arm mentioned in each publication. Details to the composition of the system in the online table (link) [38]. As can be seen, most systems use either eye tracking glasses (head-worn, 17 of 39 publications) or remote eye tracking devices (12 of 39 publications). Approaches using EOG devices are represented by 9 out of 39 publications. Robotic arms are divided into industrial robots (R: 11/39 pub.), collaborative robots (CR: 7/39 pub.), assistive robotic arms (ARA: 9/39 pub.), educational robots (EDR: 4/39 pub.), and modular prostheses (PR: 2/39 pub.). As a second characteristic, robot motion was divided into pointing to objects or positions (12/39 pub.), grasping objects (10/39 pub.), and performing tasks (17/39 pub.). Most authors chose one task, such as ADL, to proof the functionality of their system. If the system did not have an effector to grasp the object or to interact with it in some other way (e.g. with a magnet), it was reported as "point". If the object was grasped but no further interaction was reported, the classification "grasp" was chosen. Note that "task" also describes pick and place tasks, as many tasks can be accomplished by this object manipulation, such as setting a table. The type of eye movement was evaluated to determine the effects of the Midas Touch Problem or similar effects. In total, 32 of the 39 publications used fixation based approaches (F). 4 of 39 publications also used saccades. In addition, blinking (B) was detected in seven publications. The workspace in which the robot is to be controlled was divided into 2D, moving in a plane and 3D, which is necessary, for example, to place objects in shelves. Approaches of recognition 2D gaze in a 3-dimensional control of a robot is challenging. Most of the authors

presented solutions to control the robot in 3-dimensional Cartesian space (29/39 pub.). The last three characteristics describe the characteristics of the conducted user studies (28/39 pub.). Six studies included handicapped participants. In addition, Figure 3.3 presents a timeline of all included works and provides insights into the research focus, measurements, and application of artificial feedback.

Telemanipulation

Telemanipulation is usually realized by displaying digital buttons on a screen. They represent either directions and rotations of the Tool Center Point (TCP) or the robot gripper, individual joint positions, or complete tasks. In the latter case, the robot will act independently to achieve the goal of the task. In terms of the items shown in Table 3.4, most telemanipulation motions are described as "task" because complex tasks can be accomplished by manually controlling the robot using directional buttons.

The tasks used to test these telemanipulation systems were drinking [36, 48], opening doors [52], drawing [49, 50], printing cloths [47, 53], assisting with industrial tasks in shared workspaces [28], serving a meal or drink, picking up objects, changing CDs or tapes, making tea, and shaving [41].

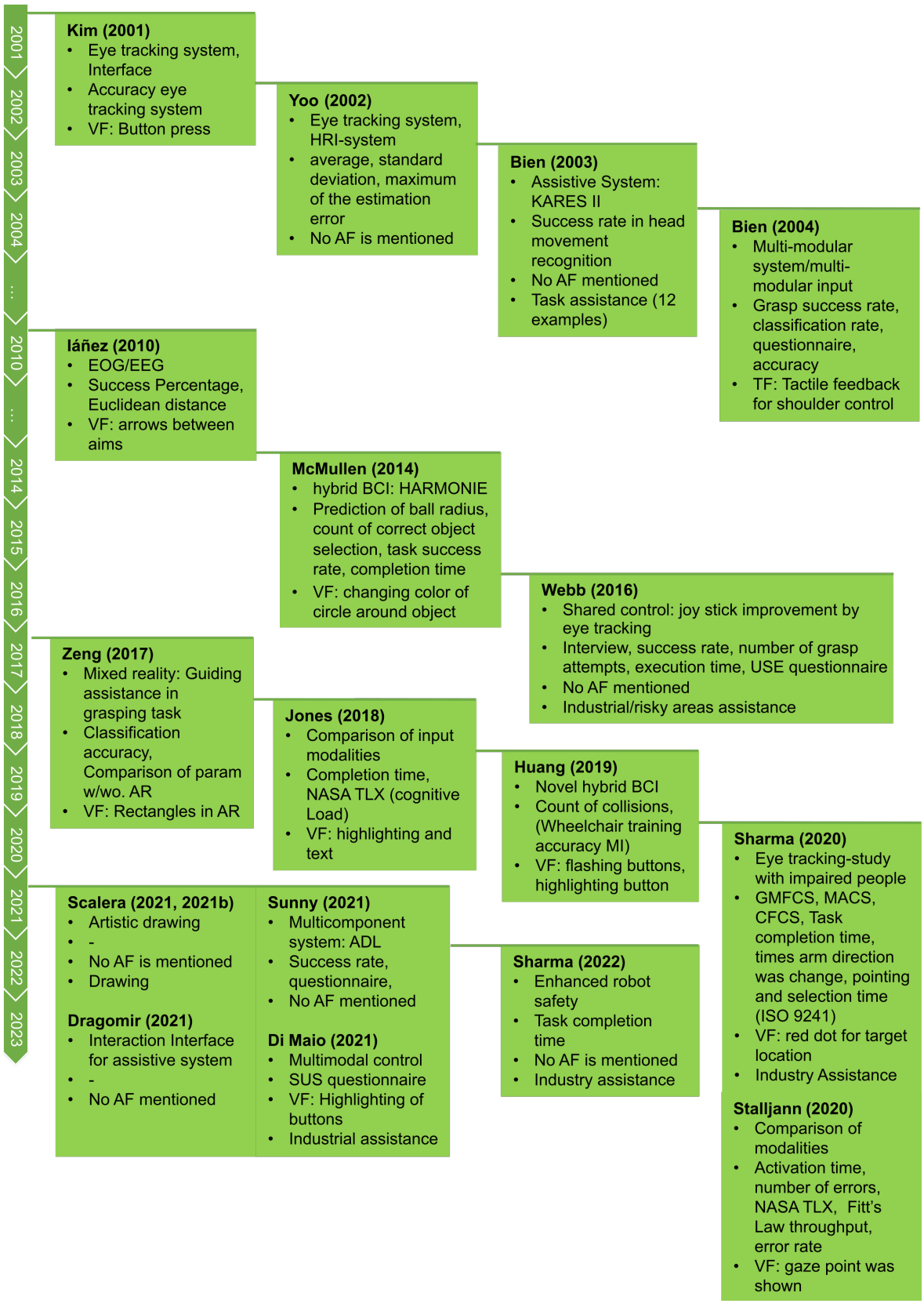
The development of eye tracking based robot control began in early 2000 (see Figure 3.3). Kim et al. [39] presented the first system based on a self-developed eye tracking system and interface to control an industrial robotic arm using on-screen buttons. Similar systems have been presented to control a robotic arm via a GUI [40, 51]. Using a GUI means that the person does not need to have the same field of view as the robot because the scene can be displayed on the screen. This is helpful for people that are bedridden. For example, an EEG/EOG-based system capable of controlling a robot from another building was presented by Iáñez et al. [42].

Although the display used in such telemanipulation approaches provides a solid basis for using either head-mounted or remote eye trackers many approaches have been found that use or comparing multimodal inputs. KARES II was one of the first systems using multiple input modalities such as eye-mouse, haptic suit, face recognition and EEG [26, 41]. For this purpose a PUMA-type robotic arm, an eye tracking setup, and a haptic suit were developed. In this work the mouth was tracked to estimate drinking intent, requiring the combination of several input modalities. The HARMONIE system demonstrates an approach using intercranial electroencephalography (iEEG) signals and eye tracking as a hBCI [43]. Eye tracking was used to move a cursor and select targets.

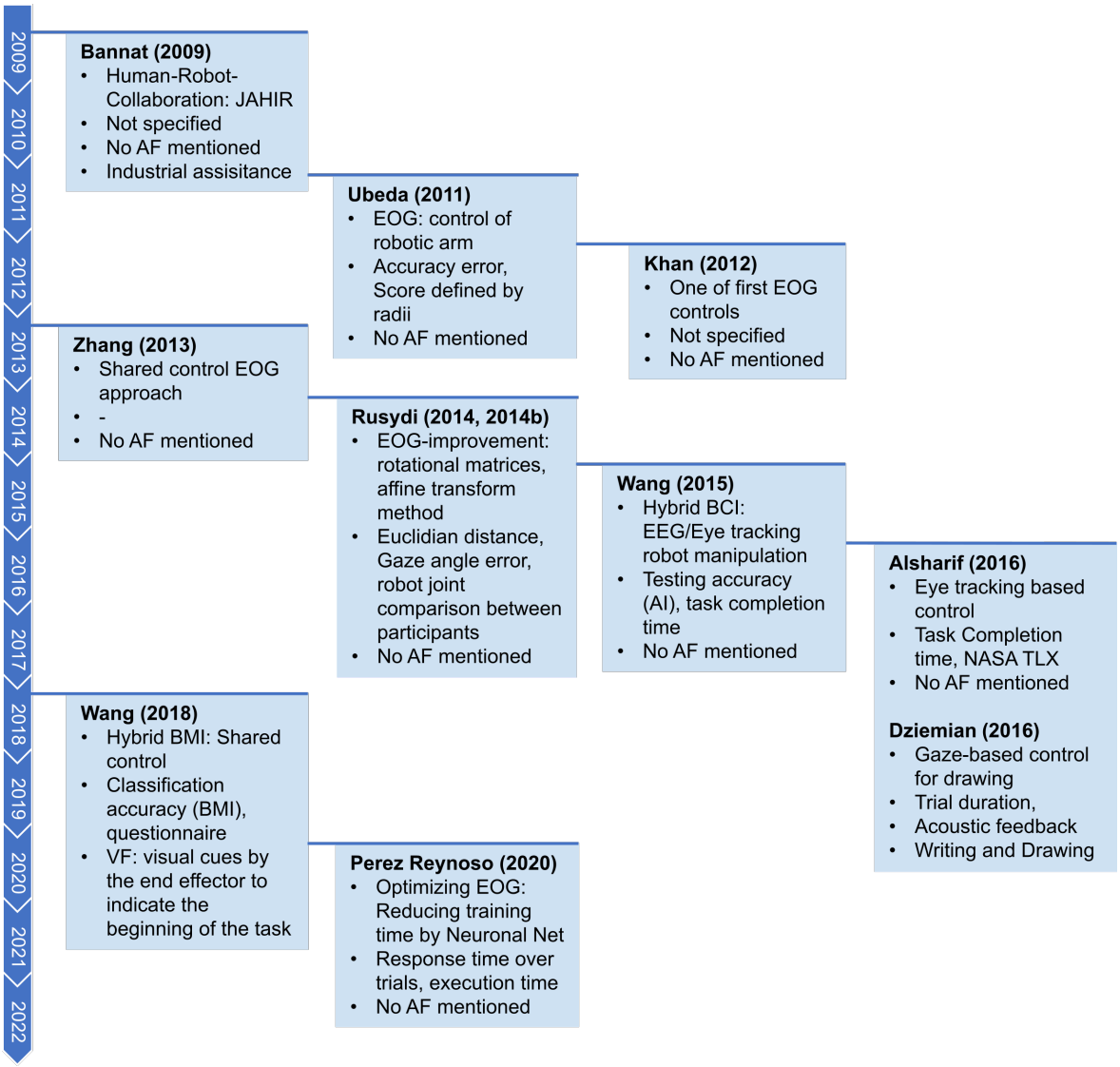
In the experiment by Jones et al. [46], the robot was used to play chess at different levels of difficulty. The two input modalities, eye tracking and joystick, were compared. Webb et al. [44] presented a system that enhanced teleoperation for a robot by using gaze to manipulate it toward a fixated object. The object was approached using a joystick controller. A comparison between head and gaze performance in a robot control task evaluated different applications between continuous and discrete control events [48]. This study demonstrated the importance of including people with disabilities, as there were measurable differences between the quadriplegic and able-bodied participants.

In several approaches, the use of multiple subsystems was targeted. In particular, the control of an electric wheelchair and an ARA, as done by Dragomir et al.[52] and Huang et al. [36], is of interest because it provides more mobility to the user. The design philosophies of such systems should include task-oriented design, "human friendliness" including safety precautions, and "modularization of subsystems" as stated by Bien et al. [41]. They combined a robotic arm, an electric wheelchair, and a mobile platform.

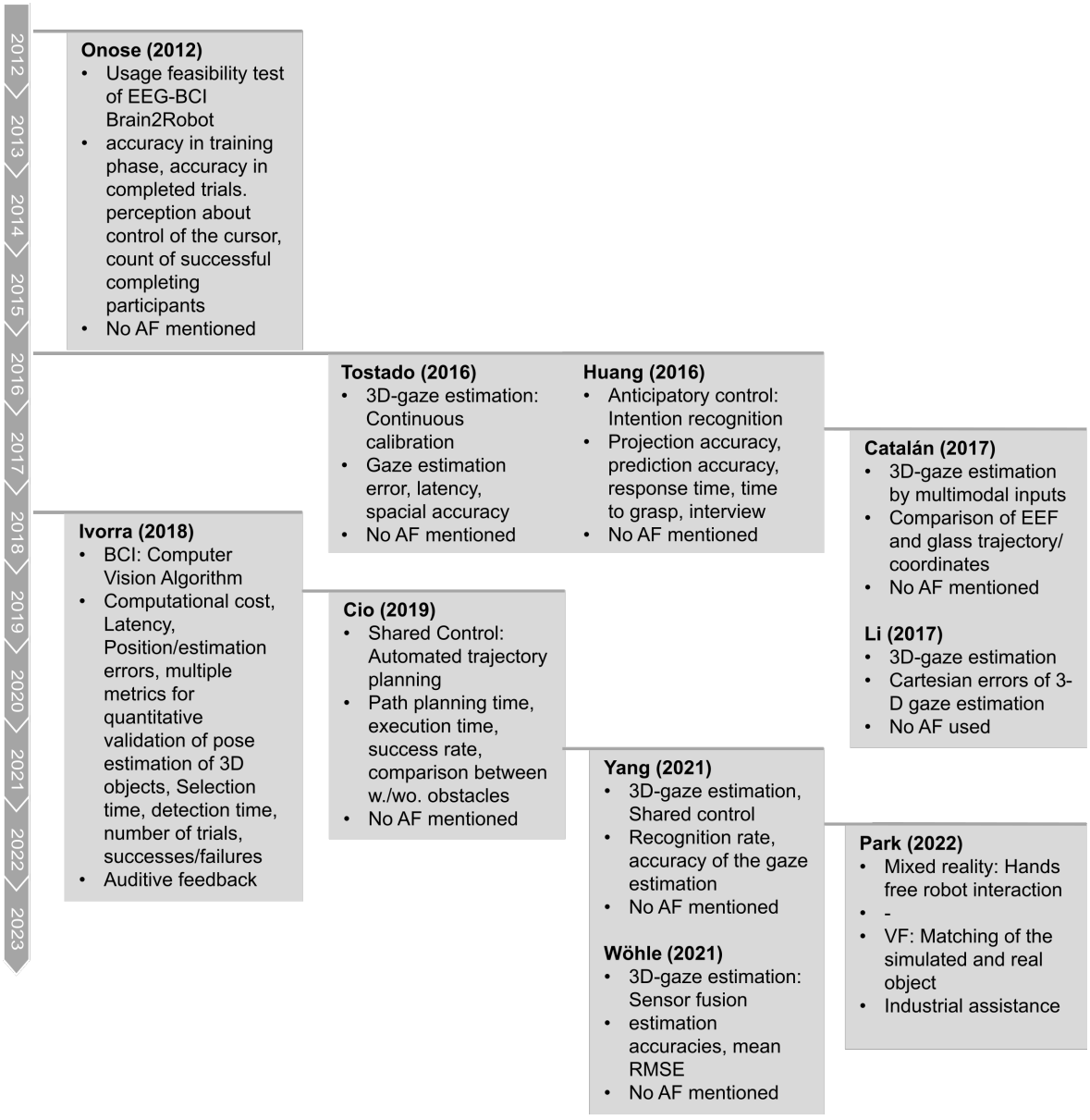
Another important advantage of the display is its ease of use for displaying visual feedback. In the approach of Zeng et al. [45], the hBCI was enhanced by using Augmented Reality (AR) inputs to correct an eye tracking pick-and-place task. In this AR environment, colored rectangles were presented as visual feedback to help sort colored objects. Others were visualizations of virtual button presses [39], tactile feedback in a multimodal shoulder control [41], and highlighting as well as text descriptions [42, 46, 43], as shown in Figure 3.3.



(a) Telemanipulation



(b) Directional gaze



(c) Object-oriented gaze

Figure 3.3: Presented characteristics in all subfigures: Research focus, evaluation parameters/metrics, use of artificial feedback, optional: use in non-assistive living environment.

Directional Gaze

Directional gaze is defined as rapid eye movements (saccades) with additional optional fixation events in a particular direction. Directional gaze can be helpful in improving the performance of the system, as quite large areas can be discriminated for a given movement. For example, Alsharif et al. [27] demonstrated one of the first systems that worked without a screen and with gaze gestures, in which participants could control a robotic arm with a specific cue of eye movements. The directional control allowed the user to perform various pick-and-place tasks. The end-effector could be translated and rotated through the interface.

Most of the grouped systems performed pick-and-place tasks [56], but this straightforward approach has also been tested with other tasks, such as writing and drawing [55, 59], according to the principle "look to the right, draw a line to the right". Other approaches let the robot perform a free trajectory in which the user visually focused on visual markers on a wall and the robot followed by interpreting EOG signals [57]. Five other publications on EOG are mentioned in this category [56, 55, 61, 62, 63]. One challenge is matching the 3D motion of the robot to the 2D motion of the eye. A solution was presented by [61] by adapting the system to user-specific parameters using a fuzzy inference system. The response time was reduced and 3-dimensional movements could be performed with 2-dimensional eye inputs by using fuzzy classifiers. With this advanced technology, the system is able to use specific coordinates to separate the signal into multiple locations in a workspace.

Other multimodal approaches, such as Wang et al. [58], combined an EEG system with an eye tracking device by implementing a trained HMM to improve the performance of the hBCI. The authors stated: "Our goal was to enable flexible and unscripted control while ensuring high reliability.". A multimodal system including gaze and EEG was used to perform a multiple obstacles grasping task. Gaze was used to correct the robot's trajectory and EEG to control the speed of the end effector [60]. Especially in shared workspaces, adding speech and buttons as input can improve accessibility in assembly tasks [54, 62].

Object-Oriented Gaze

Object-oriented gaze describes the fixation on an object to indicate an interaction to the system. This can improve the user experience because humans tend to look at the object we want to interact with. In addition, the user is not forced to look away since most of the task execution is automated and the robot is not controlled by looking in a certain direction as in the previous section. This leads to the need for computer vision to extract features from the objects as it has been applied in most cases.

Tasks such as picking ingredients with intention recognition [65], interacting with everyday objects [67, 64], or grasping a pair of scissors [20] have been realized using eye tracking techniques. The AIDE project uses EEG and EOG to move a robotic arm. The state-of-the-art algorithm uses AI to improve object selection from textureless objects in real time. Mouth poses have been used to improve user safety. The solution can be adapted to control an electric wheelchair [67].

Object-oriented gaze is a robust solution to the problem of grasping objects in 3D space. The user does not have to switch modes to move between x-y and x-z axes. 3D-gaze estimation can be used to identify the location of objects [66, 68, 17]. This information facilitates robot trajectory planning. An algorithm for continuous calibration of the eye tracking device to track the gaze as a 3D point in a scene was presented by Tostado et al. [17] and realized with a stereo vision camera and machine learning approaches. This system is usable for people with "strabism or other eye alignment defects". Magnetic, Angular Rate, and Gravity (MARG) sensors and eye tracking were fused to estimate head position in the work of Wöhle et al. [68]. Yang et al. [20] introduced a system using Apriltags to facilitate joint control with the robot. Points of interest were detected by gaze. Intention detection was used to determine whether an object should be grasped at a particular location. For 3D gaze estimation, head movements were additionally tracked. By combining artificial stereo vision and eye tracking as

an input device Cio et al. [19] showed that the system performed well in a human-guided grasping task. Even in the presence of obstacles, a task success rate of 91% was achieved. Catalán et al. [18] presented a multimodal control architecture that uses two eye tracking devices to estimate the location of objects in a scene.

In contrast to telemanipulation using a screen, an approach to the use of mixed realities has been found that also allows the use of artificial visual feedback [69]. This system has been used in industrial applications to realize a shared workspace. Through gaze selection, objects can be defined to be removed by a robot from a given space. In a service application, a system for predicting user intent in a drink mixing task was developed.

3.4 Discussion

This section discusses research questions RQ2 and RQ3. To answer these questions, a more general overview of the behavioral and technical interpretation of eye movements is presented. A comparison of input modalities, future trends, and open questions will answer RQ3 to increase transparency in this diverse field.

3.4.1 Interpreting Gaze in Robot Control

As shown, most researchers use fixation to define events. In this context, fixation is defined as focusing on an object or direction. The resulting risk of the Midas Touch Problem can be reduced by a well chosen dwell time of about 200 to 700 ms. The Midas Touch Problem is reported in several included publications [59, 27, 44, 111, 48, 82, 43, 112, 41, 46]. In some experiments, dwell time leads either to pleasant or irritating situations. It was interpreted as entertaining in the experiment of Bednarik et al. as it was seen as a challenge in a game [113] or frustrating since it caused wrong decisions in object manipulation [46]. Saccades are rarely used. One reason found is that saccades are highly noisy, unintentional, and not goal-directed, which makes them difficult to interpret [59]. Saccadic movements are used by Scalera et al. [49] to determine the trajectory of the eye movement and implement virtual brush strokes, which were then interpreted by the robot. Saccadic movements are also used to determine electrical potential changes in EOG signals [56].

Some approaches use different input modalities such as gaze gestures or blinks, although these can be highly intuitive and enrich the received gaze information [27, 114, 59, 42]. A gaze gesture is described as a sequence of eye movement elements [111] or as "the number of strokes performed in a predefined sequence" [27] and includes fixations and saccades as well as blinks and winking. If these sequences are easy to remember it may have an impact on usability.

In addition to the behavioral aspect of interpretation, the provision of interpretable data to the robot must be ensured. Not every publication provided detailed information about the software architecture. The solution was usually divided into two phases: (1) filtering and analysis of the eye-tracking data and, if used additional cameras and sensors and (2) robot trajectory planning. In the first step, OpenCV was used several times to detect the position of the pupil. Other systems like Pupil Core from Pupil Labs and the eye trackers from Tobii already provide an API to receive this data. The trajectory planning was mostly done with ROS MoveIt or GraspIt. The first ROS distribution was released in 2010. Therefore, in the first publications mentioned, ROS did not exist, so it is most likely that the robot was controlled directly by programmable logic controllers (PLC). New Python-based toolkits, modules, and environments are being developed to facilitate the control of robots and the implementation of AI models as it was presented (e.g. PyBullet, OpenVINO, Open AI Gym, and many more).

3.4.2 Benefits, Challenges and Limitations of Eye Tracking

Gaze is considered an intuitive control modality because it requires low cognitive load, is proactive, and directly correlates with action intentions [17]. It provides natural, effortless, and rich information that can be interpreted by the robot [66, 45], especially since humans tend to look at objects they want to interact with [66]. When designing interfaces for assistive devices, it is important to consider people's ability to retrain their motor functions, which leads to different approaches solving a challenge. During the training phase, the setup may be perceived as unnatural and distracting, resulting in additional cognitive load [17]. This may change over time. Considering that the system will be worn for several hours in real-world applications, such as head-mounted eye tracking devices, the system should be lightweight and adjustable to reduce discomfort and pressure on the head [68]. In terms of social factors and the individuality of each user, there are important influences that challenge head-mounted and remote eye trackers. Parts of the environment are usually included in the scene video recorded by the eye tracking camera, which shows limitations regarding privacy regulations [68]. As stated by the European Data Protection Board, the increased use of smart cameras today leads to a large generation of additional data to the captured video itself. This leads to an increased risk of secondary use for unexpected purposes such as marketing and performance monitoring [115]. Especially in combination with the use of the robot in ADL, such as for maintaining personal hygiene, ways must be provided to avoid the monitoring of such sensitive data. According to a 2020 report, more than half of humanity has either myopia or presbyopia [116], requiring corrective devices such as contact lenses or glasses. Most eye tracking devices are not usable with glasses or lead to higher inaccuracies [7, 107].

Inadequate accuracy is often reported for eye tracking devices [46, 48, 65]. The signal-to-noise ratio (SNR) in EOG and VOG is sometimes insufficient [107, 62, 60, 59, 36]. This makes it difficult for users to complete the task and limits bandwidth, resulting in incorrect button and object selection [65, 60, 59]. In VOG and IOG applications, head movements can lead to moving areas of interest [7] due to limited performance of gaze-mapping algorithms [20]. In experiments, this results in a limitation of head motion [20]. Other limitations include lighting conditions, camera field of view, and object overlap [105]. Eye tracking glasses should be calibrated regularly or secured with a strap, as slippage can lead to errors [48]. A common problem with IMUs or gyroscopes built into the glasses is DC offset. In gyroscopes, magnetic materials can also cause data drift. They are often used to track head or body movements [68, 107]. EOG signals must be filtered and segmented to obtain the desired eye movements for control, as unintended eye movements are also recorded [36]. Gaze directions are difficult to track in previous work [36], but solutions are being found in current work [61]. Information is lost through filtering [42]. One reason why EOG signals are easy to use is the linear relationship between signal and eye movement displacement [63].

Comparison of Input Modalities

The decision of which input modality is more appropriate depends strongly on the technology used, the goal of the task, and the control algorithm. A number of comparative studies with different experimental setups were identified in this review [46, 48, 107, 109, 108]. Figure 3.4 shows the articles and the input modalities. The task description is given to provide a deeper insight into the study design.

Overall, VOG and IOG were rated more positively than the other modalities. Positive aspects were identified as subjective rating (low workload/cognitive load [109, 48], comfort [107], and robust functionality [107]. Pasqualotto et al. [109] suggested its use as a communication device, similar to Schäfer et al. [107], who stated its use for discrete events, such as trigger events, based on the traceability of fast eye movements. Negative aspects have been reported in poor performance with the Fitt's law test, which is associated with low accuracy [108, 46]. It describes the relationship between the time it takes to move quickly to a target area and the size of that target [108].

Contrasting results have been found with other modalities such as the thumbstick [46] and the

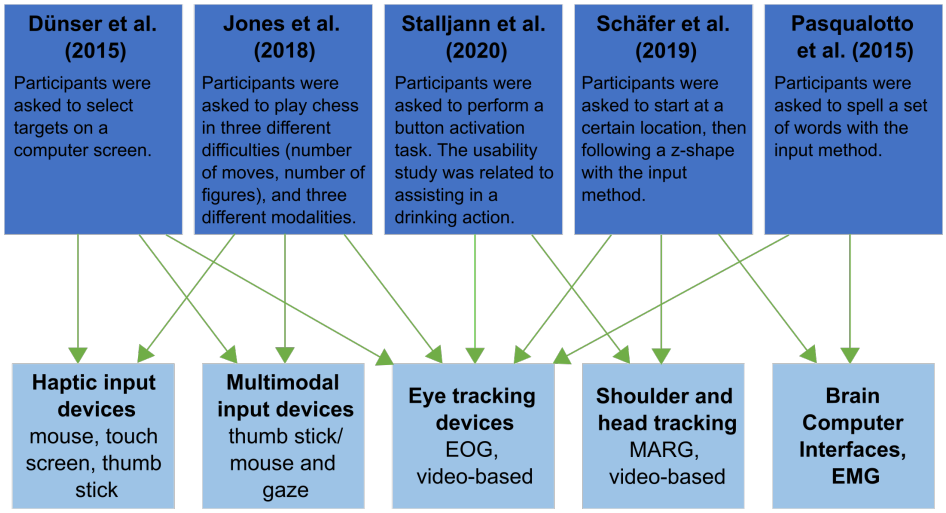


Figure 3.4: Identified comparative studies describing the task and the used input modalities.

mouse [108]. Jones et al. [46] found better performance when combining input modalities than when using eye tracking alone, as it was faster and found better results in workload. However, Dünser et al. [108] reported worse results for the eye tracking approach than for the mouse and touchscreen in both objective and subjective measures. One reason for this may be that most participants have used one of these before [40, 46]. It is difficult to determine whether the differences in parameters are due to learning effect or pure perception.

Head tracking using a Magnetic, Angular Rate, and Gravity (MARG) sensor has also been positively evaluated [48],[107]. This type of sensor can provide comprehensive data on humans movement. Stalljann et al. [48] presented that button activation to perform the task was less demanding, strenuous and frustrating with this sensor. Shoulder or head control is considered less intuitive than hand control because the information density is reduced. It is also mentioned that it can be more natural than other input devices [26]. In comparison, shoulder control was found to be a suitable candidate for moving a robotic arm in continuous control and was rated better than eye tracking in one case [48]. The patient with tetraplegia was able to perform better with eye tracking as measured by the lower error rate. The difference to the MARG sensor was not significant in the group of able-bodied participants.

EEG, EOG, and BCI were rated lower than all other input modalities due to low signal transmission [107], additional hardware required [107], and more effort and time required to perform the task [109].

3.4.3 Future Trends

The result of research question RQ3 showed an increased mention of the following topics: multimodal control, including 3D gaze estimation and hBCI, and shared control. The authors see high potential in these topics as they can reduce the error rate of eye tracking devices used as stand-alone systems, improve the usability of such systems, and address individual needs.

Multimodal Control Inputs

Multimodal control inputs can improve the user experience for each participant. They are combined to adapt to the degree and type of impairment [41, 110] and to the task [28, 54], increasing the intuitive control of the complex system [54, 66, 53]. The system of Bien et al. [41] presents a hierarchy and design for different levels of disability. The user was able to move the system with the shoulder if the upper body motor functions were still intact. If the patient has a progressive disease, he can later

switch to eye tracking control. The use of multiple input modalities leads to improved performance [105], such as combining gaze and joystick control [44] or gaze and head tracking [39].

In this work, 6 of the 39 identified and included papers involved the use of hBCI that included eye tracking devices. Eye tracking was used for rapid intention detection, such as moving the end-effector in a particular direction [58, 60], for object pose estimation [64, 67], or for triggering switch events [45]. For detailed information, we refer to [105, 117, 118, 109, 119].

3D gaze estimation can be achieved by combining different input modalities. By estimating the intersection of the lines of sight, we can determine the 3D gaze position with respect to the scene camera in head-mounted glasses [20, 17]. Due to head movements, the point on a world coordinate system needed for robot manipulation is not stable. Two approaches that combine head tracking with eye tracking have been identified. The first approach used motion detection from video cameras placed around the user [64, 20, 68]. The second approach uses accelerometers, IMUs and MARG sensors placed on the user's neck and head [68]. The latter is mostly used as a control based on head motion unrelated to 3D gaze estimation [48, 107].

Shared Control

A synonym for shared control was found to be described as semi-autonomous robot control. The robot can act and perform movements or tasks automatically, but is dependent on a user. The goal is to realize a comfortable and intuitive control of the device [54, 66, 53]. One solution is to incorporate intention recognition into the design process [65, 98, 41]. Intention can be read by Areas of Interest (AOI) in eye tracking approaches, such as the handle of a cup when the user intends to grasp it or by tracking facial features. Further literature on this area of research can be found in [89, 91, 120], and [121].

Task-oriented design is one of the key benefits of shared control. Activities of daily living are composed of multiple subtasks. For example drinking can be divided into reaching for and grasping a bottle, filling the glass, placing the bottle, taking the glass and reaching toward the user. A human will decide at which step to stop, and it is not necessary to perform them all in one go. In most of the presented approaches, the task is realized by the user fixating the object and correcting the system [58, 65, 43, 26, 41]. A comparison of the resulting degree of autonomy is shown for a feeding task by Bhattacharjee et al.[15]. The main finding is that the user should always feel in control of the situation and control.

One way to implement shared control is to use of AI, as has been done in several works. Detecting of an object can be easily done using models such as YOLO or RetinaNet [67, 69]. Neural networks have already been implemented to learn robot motion [41], to model EOG functions [61], to model this eye-hand coordination behavior during grasping [66], or for face recognition [53]. The use of AI helps to automate tasks. Especially in robotics, but also in behavioral science, these models are becoming increasingly important.

3.4.4 Open Questions

To conclude RQ3, open questions were uncovered that will be necessary progressing in the design process of the eye tracking control system for robotic arms. Frequently mentioned research topics focused on performance measurement, inclusion of people with disabilities in studies, and the use of artificial feedback to improve the usability and accessibility of the system.

Assistive Robotics Performance Measurements

How the performance of an ARA is measured is highly dependent on the system, the domain of origin (e.g. computer science), and the study design. It was found that the most common parameters reported in the included papers were task completion time and success rate (see link). These parameters are not

comparable between studies because of differences in design. All reported parameters were divided into task-related parameters, computational parameters and empirical parameters. Computational parameters are mostly measured as accuracy [66, 41, 20, 64, 58, 45, 36]. In the case of object recognition, parameters such as prediction accuracy, projection accuracy [65, 43, 67], recognition rate [20, 57], and reaction time [65] are used. Computational cost, as reported by Ivorra et al. [67], was used as a way to rank different algorithms. Task-related parameters are most often given as task completion time and success rate. Counting different events also helps the reader to estimate the challenges in the system. For example, authors mention the number of required gestures, which indicates the effort required by the user [27], or the number of collisions when there are obstacles [36]. In the case of eye tracking, gaze parameters such as dwell time (fixation duration), gaze estimation accuracy, or areas of interest are controlled [7, 31]. Other errors have been reported by authors such as Euclidean and Cartesian errors to describe the drift from the intended position to the robot position [66, 68, 17, 61]. We want to show how to ensure better comparability between studies by specifying exemplary parameters. The completion time itself depends on several parameters. Dividing the task into several steps has been found to be helpful, such as different empirical and computational parameters, (e.g. reaction time of the user and the computation and the execution time with the robot). Reproducible results can be obtained by additionally specifying the trajectory length and the maximum speed of the robot. Boundary conditions for reproducibility were missing in some publications.

User satisfaction is evaluated through subjective measurements. The following tests and questionnaires have been identified: NASA TLX (NASA Task Load Index), USE questionnaire (Usability, Satisfaction and Ease of use), SUS questionnaire (System Usability Scale), and QUEST 2.0 (Quebec User Evaluation of Satisfaction with assistive Technology). Non-standardized tests were often used to analyze specific features of the system. This section has been expanded to include the QUEAD (Questionnaire for the Evaluation of Physical Assistive Devices), based on the TAM (Technology Acceptance Model), to provide another test method applicable to assistive robotics. Table 3.5 should provide a quick overview for new researchers. The question to be answered in future work is to what extent these questionnaires can be used to evaluate performance in an eye tracking based robot control. Although most of the questionnaires are well established in science and hurdles in execution and evaluation have been uncovered, the novelty and therefore the application to this topic needs to be explored.

User-centered Design

In this review, six of the 39 publications included people with physically disabilities. A lack of inclusion of disabled people in such studies can be seen in other areas, such as the accessibility of wearables [110]. Participants' opinions about the system are crucial for improving such systems. Differences in the evaluation between able-bodied and disabled participants have been reported several times [48, 109, 41, 94]. For example, Stalljann et al. [48] found that healthy participants could easily switch between controls and interpret the artificial feedback well. This was much more difficult for the tetraplegic participant. Onose et al. [64] cited difficulties caused by changing postures in a multi-day study. In particular, severely immobile users may find it difficult to remain in the field of view of stationary eye tracking devices [43].

According to Lulé et al. [130], the decision against life-prolonging treatments in ALS patients is based on the fear of loss of autonomy caused by lack of mobility and aggravated communication. This has a mayor impact on the Quality of Life (QoL). Eye tracking based controls can help to increase the possibilities of an autonomous life [7]. Moreover, everyday use of such systems poses risks and challenges. Regarding an ARA mounted on an electric wheelchair, the needed instruments of a tetraplegic person can include respiratory and gastric catheters as well as communication devices [7].

Therefore, the movement of the robot has to be done carefully in close proximity to the person so

Table 3.5: Standardized questionnaires for subjective evaluation of assistive robotics.

Questionnaire	Applicable for the evaluation of	Number of Items	Used in included works	Limitations
NASA TLX [122]	Task-dependent human-centered evaluation of workload	6 items, 100 or 50 point Likert scale	[27, 48]	Failed test of construct validation and varying interpretation between participants [123]
USE [124]	Measurement of the subjective usability of a product or a service [125]	4 categories with a total of 30 items, 7 point Likert rating scale	[66, 44]	-. Refinement due to validation enhancement was presented by [125]
SUS [126]	Global assessment of system usability	10 items, 5- or 7-point Likert scale	[28]	Usability cannot be measured in an absolute way [126]. For a detailed insight in the subjective dimensions efficiency, effectiveness and satisfaction other questionnaires should be conducted.
QUEST 2.0 [127]	Satisfaction of an assistive device or service	12 items in two dimensions (Device and Service), 5-point satisfaction rating scale	[20]	Generic Assessment, items related to certain Assistive devices may be absent, yet can be added. [128]
QUEAD [129]	"Subjective perception of usability and acceptance of a new physical assistance system or control mode." - [129]	5 categories with a total of 26 items, some categories can be used separately. 7- or 5-point Likert scale. Tested on an assistive robot	-	Till now no QUEAD scale benchmark
Non standardized questionnaires	Varying content depending of the researchers interest	-	[51, 65, 64, 46, 48]	Biasing the participants should be avoided. Significance etc. can vary.

as not to interfere with the tubes and cables. In addition to medical challenges, the research in this area should be guided on what activities and needs should be addressed and are necessary to provide an improvement in QoL for each individual. The performance of ADL with the use of ARA may be a promising solution to improve QoL. Care is mostly provided by family or professional caregivers. They take care of personal hygiene, dressing, medication, and in the case of assistive devices such as the ARA, maintenance and daily setup of the system [131]. Especially when the patient has invasive devices, a clean environment is necessary. The system should be easy to keep sterile. In addition, the system should be easy to deploy to further reduce time pressures on caregivers.

Based on these impressions, the question arises as to what other effects have not yet been found in the case studies. Further studies need to be conducted with real-world robots controlled by eye tracking to assess the effects on impaired and able-bodied people. In these systems, multiple users are involved in creating a user-friendly design. The design steps should include additional input from caregivers and family members to ensure ease of use. Due to the novelty of the technology no details could be found.

Artificial Feedback

Artificial feedback is well studied in the field of electric prostheses or other computer-based user interfaces. It helps to provide an easy-to-interpret signal for closed-loop interaction between the user and the robot, and provides a good user experience when applied correctly. The question arises as to how such feedback can be used in robotic applications and eye tracking. Figure 3.3 shows the use of artificial feedback in the form of auditory, visual, and tactile feedback in the included publications. Summarizing all included publications, 25 authors reported no use of feedback. 11 used visual feedback, two used auditory feedback and in one publication was tactile feedback used.

In the telemanipulation systems group, half of the authors used visual feedback to indicate system performance to the user. In most APIs, button click visualization, colored boxes, outlines, and bounding boxes are easy to program and provide visual feedback to the user about their interaction with the system. Transferring these feedback methods to a controller without a display seems complicated and may be unnecessary if the robot responds quickly. This may be one reason why most studies have not suggested the need for further artificial feedback. However, it can provide important information to the user, resulting in higher user satisfaction [45]. In the identified publications, two authors used visual feedback without the use of a display. Wang et al. [60] used the robotic arm itself to provide visual cues by moving the end-effector to trigger each task. Park et al. [69] used virtual objects displayed in augmented reality to visualize the movement of the object, which was then performed by the robot. The use of virtual and augmented reality allows this discipline to easily integrate visual ways to easily integrate visual feedback into real-world tasks.

Auditive feedback can be easily applied when using a tablet [7]. Examples of the use of auditory feedback were provided by Ivorra et al. [67] and [59]. In these papers, a brief tone indicated either the selection of an object [67] or the successful selection of a movement command [59]. Design rules for auditory feedback can be found in the work of Tajadura-Jiménez et al. [132].

Tactile feedback can be used. With small vibration motors like those used in smartphones. Vibrotactile feedback can be realized, which could be mounted on the eye tracking glasses [133, 134]. Bien et al. [41] used it to provide status information to the user. There are several design rules for vibrotactile feedback that should be considered [135, 136, 137]. These are mostly related to either body-worn devices or integrated into eye tracking glasses. While this is a good start, more research needs to be done on what feedback is practical and whether it affects user or system behavior.

3.4.5 Limitations

Interest in assistive devices will grow in the coming years due to the needs of the elderly and disabled, as well as the decline in caregivers. Each of the research interests presented above is a vast domain in

itself. Therefore, a bias based on the limitation of robotic application cannot be excluded. To reduce this bias and to enrich the discussion, the excluded works were partially included. The focus was on research that considered eye tracking in combination with a robotic arm. Due to the interdisciplinary nature of the field, different results for advantages and disadvantages may be found in literature focusing on only one of the topics.

The bias was estimated by comparing the keywords given by the authors and the search terms. Figure 3.5 shows the results for the keywords listed in section 3.2. The top 10 keywords used in all publications are 1. robot, 2. human, 3. interface, 4. control, 5. gaze, 6. robotic, 7. eye, 8. interaction, 9. assistive, 10. computer. The reason why 17 relevant publications were found in the snowball search is based on the absence of keywords such as "assistive", "impairment" and "shared" and the fact that the database was not included in the initial search. Regarding the search terms used in the methods, "tracking" was not among the top 10 keywords in the list of systematically searched papers. The keywords "brain", "machine", "BCI" and "EOG" were identified in the snowball search. Since the focus of this paper is on eye tracking based systems, the identification and screening process was not repeated. Regarding the top 10 list of all publications, a new search would not have yielded any new results because most of the terms are synonyms of a previous used search term or would have missed key terms.

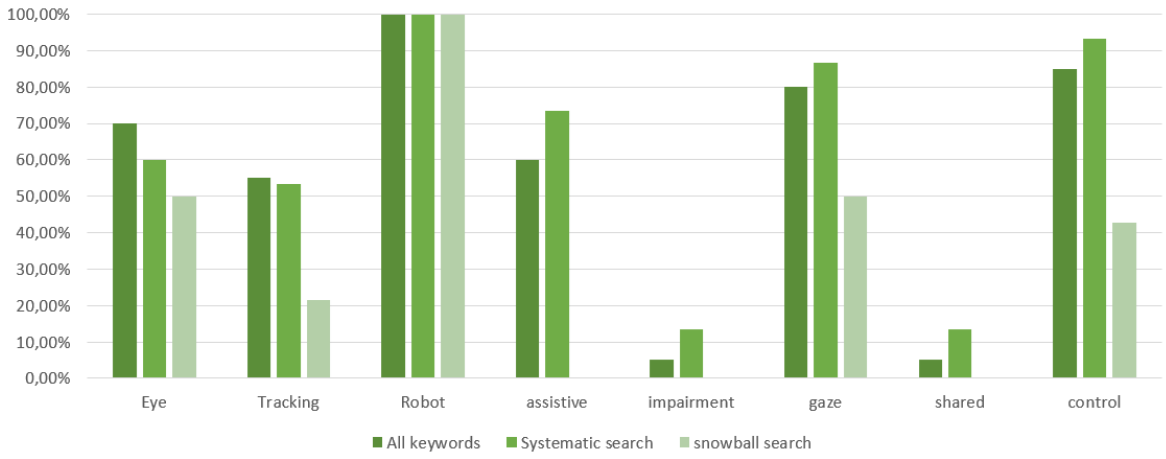


Figure 3.5: Search term analysis.

3.5 Conclusion

We presented an overview of 39 works with real robotic applications controlled by eye tracking. These approaches require interdisciplinary knowledge from the fields of robotics, human-robot interaction, and behavioral science. The interdisciplinary nature of the field often leads to different keywords for the same topic, making it difficult to find literature. It has been shown that the interpretation of gaze depends on the hardware used, the task the robot has to perform, and the and the stage of the user's disability. Research interests were found in topics such as multimodal inputs and shared control. Insight was provided on performance measurements, inclusion of impaired people in study design and artificial feedback. Benefits and challenges in terms of eye movement detection methods and various important user parameters were decribed. The comparison of input modalities showed that the optimal devices depend strongly on the user's abilities. The table presented online (link and link) will be expanded to further facilitate the identification of assistive robotic arm controllers.

Chapter 4

Publication II: Requirements of people with disabilities and caregivers for robotics: A case study

Summary: Due to a collaboration with Lebenshilfe Kinzig- und Elztal e.V., I gained profound insight into the daily lives of people with disabilities. As the scoping review in Chapter 3 revealed, including people with disabilities in research improves the design process through their assessments of accessibility and usability. The literature review for this study revealed different robotics requirements between facilities in different countries. In Germany, strict privacy regulations govern the use of camera systems, including those in robotic systems. The application of robotics has proven beneficial in many cases by reducing caregivers' workload and increasing users' independence. This requirements analysis investigated the ethical restrictions and the abilities of potential users in a German assisted living facility and informed the design of the eye-tracking-driven control presented in Chapter 7.

This study examined the requirements for deploying a social robot in a shared apartment to assist with daily living activities. Although social robots differ greatly from robotic arms and assist users in different ways, the lessons learned regarding the usage of private data can be applied to various robot applications. During the initial interview with care management and staff, we discussed the ideal robotic system and its capabilities for assisting with everyday tasks. This led to a list of tasks that the robot should be able to perform.

Ethical considerations were kept in mind when designing the control. Eye-tracking glasses are preferable to stationary eye-tracking devices because they do not primarily film the person's body and face, thus reassuring the initial design choices presented in Chapter 5.

Contribution: Anke Fischer-Janzen was the primary author of this publication. She was responsible for its conception, implementation, evaluation, and manuscript preparation. Markus Gapp was responsible for implementing and conducting the study revolving around assistive work. He contributed to the conception, methodological discussions, and manuscript review. Jennifer J. Blöchle and Katrin-Misel Pronomarjova contributed to the experimental design and execution. Marcus Götten and Thomas Bartscherer contributed to the manuscript review and provided technical support. Thomas M. Wendt and Kristof Van Laerhoven contributed through supervision, methodological discussions, and manuscript review.

Requirements of people with disabilities and caregivers for robotics: A case study

Anke Fischer-Janzen
Offenburg University of Applied Sciences

Markus Gapp
Offenburg University of Applied Sciences

Marcus Götten
Trier University of Applied Sciences

Katrin-Misel Ponomarjova
Offenburg University of Applied Sciences

Jennifer J. Blöchle
Offenburg University of Applied Sciences

Thomas M. Wendt
Offenburg University of Applied Sciences

Kristof Van Laerhoven
University of Siegen

Thomas Bartscherer
Trier University of Applied Sciences

Abstract

Robotics offers new solutions for digital customer interaction. Social robots can be used in applications such as customer support, guiding people to a location on company premises, or entertainment and education. An emerging area of research is the application in community facilities for people with disabilities. Such facilities face a shortage of skilled workers that could be addressed by robotics. In this work, the application of social and collaborative robots in care facilities and workshops for the disabled is presented by providing a requirements analysis. The use of the humanoid robot Pepper in assisted living was tested and subsequently evaluated in interviews with caregivers who initiated and observed the interaction between the group and the robot. Additionally, robotic applications in assisted work were assessed, resulting in a divergence from the industrial use of robots. A comparative overview with recent literature is presented. The connection between the community home and the workshop raised the question of whether the use of different robots in both places could lead to conflicts.

4.1 Introduction

An ongoing trend is the use of robotics in social and medical environments. Early versions of social robots are represented in the early 2000s with toys such as Furby (Tiger Electronics) or Aibo (Sony) that change their behavior depending on the user's behavior [138]. Today, social robots are motivated by new challenges such as the prevention of loneliness and social isolation as seen in the Covid-19 pandemic [86],[139], assistance in activities of daily living (ADL) through cognitive and physical stimulation [86], and human-friendly navigation to avoid collisions in logistic tasks, but also in facility tours to present the location to customers [140]. In healthcare, they are used to assist healthcare workers by reducing time-consuming tasks such as picking up and delivering bed sheets and other supplies, allowing caregivers to focus on patient needs and more critical tasks [2]. As a result, the market and research interest in social robots is growing [141],[87].

The application of social robots varies greatly depending on the audience and purpose. According to the IFR (International Federation of Robotics), they are categorized as a subgroup of service robotics, as opposed to industrial robotics [142]. Service robots are designed to perform human-centered tasks outside of industrial automation applications [143]. A division into social and non-social robots and domestic or professional use further defines the application of the robot [144]. For this reason, design focuses not only on functionality, but also on form and material to increase acceptance [86]. This can be seen in the diverse field of social robotics, from humanoid robots such as Pepper and NAO (Aldebaran United Robotics Groups) [2], to pet-like companion robots like Spot (Boston Dynamics) and PARO (PARORobots USA), to service robots used in health and social care facilities to optimize logistics and reduce time pressure on caregivers such as Moxi Robot ([2]) and Relay (Swisslog Healthcare).

In contrast, assistive robots, as they are colloquially known, are used for rehabilitation or assistance in everyday life. They do not have the ability to speak and are mostly based on robotic arms, but still need adaptive capabilities. Rehabilitation robots have the ability to guide the user's extremities to assist in the rehabilitation process [2]. The robots' force feedback allows progress to be tracked and adapted to each individual patient. The gamification of such processes is being explored through the implementation of computer games or virtual reality [2]. Assistive Robotic Arms (ARA) can be attached to wheelchairs to assist in performing daily tasks through a joystick, head tracking, eye tracking, or brain computer interface (BCI). The control of such systems can be facilitated by shared control that automates task execution for example by using intention recognition [20].

With the ability to recognize emotions and personality and store memories of previous interactions, such robotic systems can benefit care facilities and hospitals [141]. Adaptive behavioral models, cognitive architectures design, and the ability for perceived empathy in human machine interaction are recent challenges. The application of artificial intelligence accelerates the solution finding [87].

These challenges and requirements must be evaluated separately for each individual robot and target group (e.g., customers, elderly people or people with disabilities). In this work, the user requirements have been identified in cooperation with two related organizations. A case study is presented in which requirements and solutions for assisted work were identified and the robot Pepper was evaluated in an assisted living facility. A comparative overview with existing literature was performed to validate the results.

4.2 Methods

This work is based on a cooperation with Lebenshilfe im Kinzig- und Elztal e.V. and the company WfB Haslach gGmbH. Lebenshilfe im Kinzig- und Elztal e.V. is an organization that provides services for people with disabilities to ensure participation in all areas of life and to maintain a good quality of life. Services include counseling for family members, assistance in finding housing, suitable employment, and leisure activities [145].Lebenshilfe im Kinzig- und Elztal e.V. is the shareholder of WfB Haslach gGmbH. This company offers people with disabilities a wide range of work and employment opportunities that are adapted to each individual's abilities [146].

In a multi-step process, as shown in Figure 1, the authors identified the needs of both institutions with respect to robotic applications. This work took place between March and October 2023. After an on-site analysis, the evaluation was split into two requirements analyses due to the different needs and work areas. While in the assisted work a demand was found to relieve the caregivers by reducing quality control in production, in assisted living entertainment and assistance in cleaning and logistics were identified, which will be presented in detail in the following sections.

The robot Pepper was used to evaluate the user experience in assisted living (Figure 4.3). The robot is 1.20 m tall and weighs 28 kg. The battery life is 12 hours according to the manufacturer. Pepper was equipped with applications such as games (memory, guessing, and animal imitation) and interaction options. The interactions were limited to hugging people, speaking fixed scripts, posing for selfies, and recording greetings to relatives or friends. The robot is able to locate itself and move to predetermined locations. A speech recognition tool provides the robot with an additional input modality to the touch screen.

4.3 Results

The analysis began with a visit to both facilities. An on-site evaluation of both facilities was conducted. The authors spoke with management and staff. A standardized questionnaire was not used. The requirements of each facility are presented below.

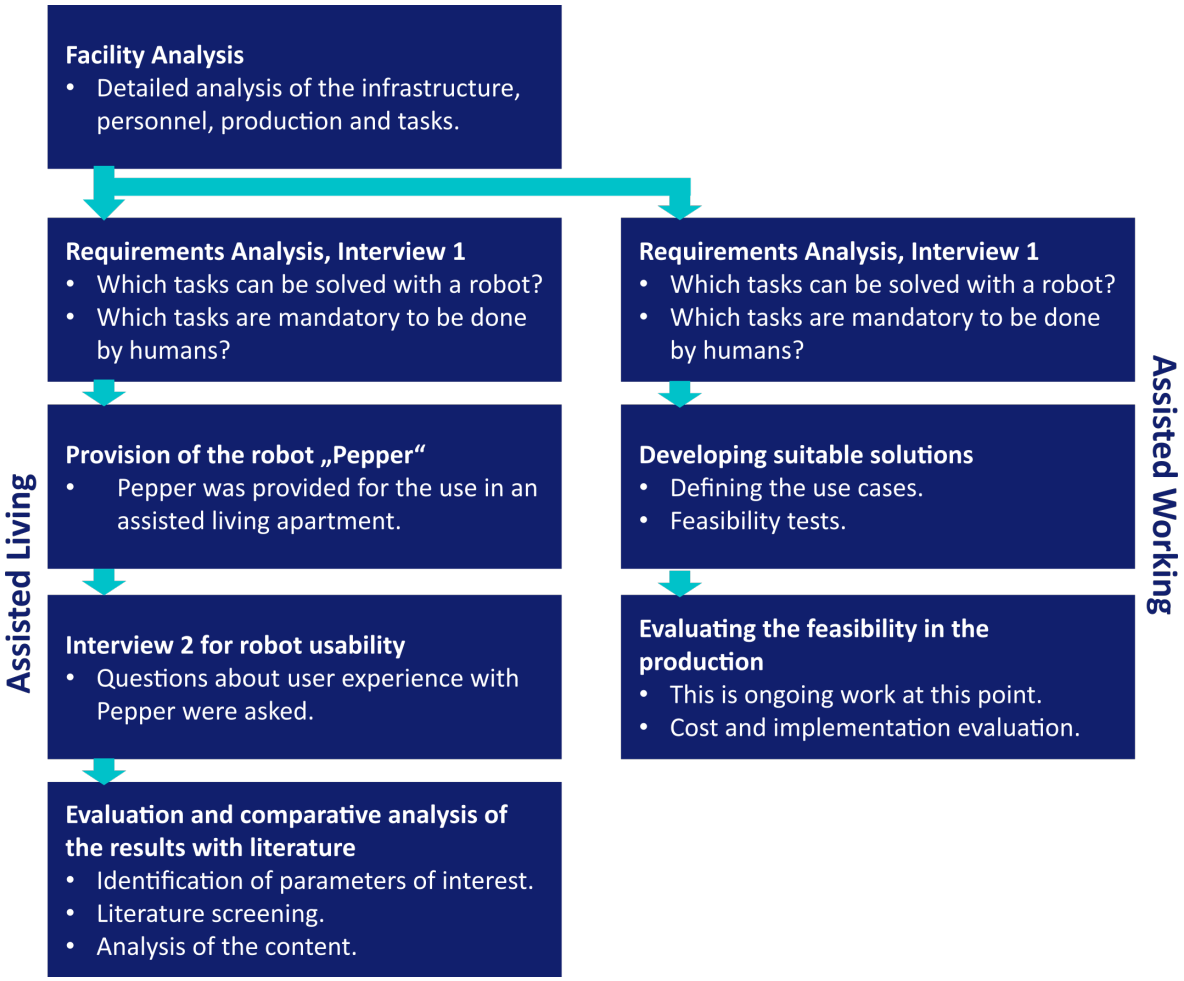


Figure 4.1: Application and requirement analysis process.

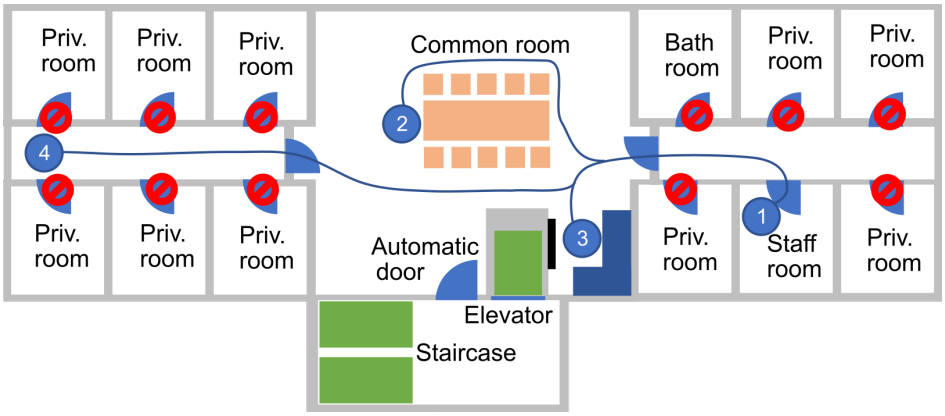


Figure 4.2: Example of the shared apartment. It is accessible through an automatic door that can be opened by pressing a button. 1. Robot is loaded in this place overnight. 2. and 3. Robot location for entertainment interaction. 4. Trajectory to guide residents back to their rooms. Private rooms and the bathroom should not be accessed by the robot.

4.3.1 Assisted Living

The community housing complex visited is divided into three communal apartments, each with space for 10 residents (Figure 4.2). The residents are adults and pensioners. Most of the residents work at WfB Haslach gGmbH. People with different disabilities live together. They are cared for around the clock by experienced caregivers, but they do most of the daily tasks themselves, such as cooking together and keeping their rooms clean. To ensure equality for people with disabilities the German government supports housings like this. One goal is to allow people with disabilities to choose the type of housing that best suits their abilities and preferences. Another goal is to decentralize the complexes, which will lead to less segregated and closed living environments [147]. The modern complex represents this agenda. Since every part of the house is accessible, it provides perfect conditions for social, mobile robots such as the humanoid robot Pepper.

The requirement analysis was conducted in the steps mentioned in Figure 4.1. First, caregivers and managers were asked, what features they thought the optimal robot should have and what tasks the robot would be most helpful with. The results showed that the robot would be most helpful if it could guide people to their rooms, retrieve objects from other rooms, such as a glass of water from the kitchen, assist with eating, or provide general entertainment functions such as playing a game. Due to the Covid-19 pandemic, a feature that regularly cleans surfaces and door handles according to the assisted living standards would be helpful. When walking, some residents may need assistance due to the risk of falling, or they may walk slower than the rest of the group, in which the robot could overtake the role of an additional caregiver to walk with them at their own pace. On the other hand, the robot should not access areas such as the bathroom due to privacy restrictions, nor should it distribute medication. Although the robot could help by monitoring and assisting in cleaning each resident's room, the management pointed out that this would raise further ethical questions about privacy regulations. Due to the infrastructure of the building, it was also requested that the robot could be able to ride the elevator or go outside the building to assist with walking activities.

In a second step, the organization was interested in testing such a robot in the shared apartment. The Pepper robot was assigned to one floor of the building, but the residents could visit each other as the rooms are connected to a common room on each floor. Over a period of eleven days, the robot was active for at least one hour every day except for two days. Caregivers monitored the interactions and ensured a safe and privacy-compliant environment for the residents. Therefore, the robot was only used in the common areas.

Table 4.1: Feedback key points regarding the usage of Pepper.

No.	Key point	Perception
1	Mood-lifting effect.	positive
2	Speech abilities: discreet with personal information.	positive
3	Setup time of about two hours.	positive
4	Limited localization accuracy.	negative
5	No artificial feedback to signal the end of a task.	negative
6	Not usable in a loud group.	negative
7	Switched languages unintentionally.	negative
8	No pressure sensors to detect contact.	negative
9	Permanent upper body movements.	negative
10	Maximum volume setting was too low.	negative
11	Proposal for a more stable standing position.	-
12	Proposal for smart home device control.	-
13	Speech model and hardware was perceived as outdated.	negative
14	Proposal that robot should be able to play music.	-

In the third step, the caregivers collected the residents' experiences with Pepper and participated in the second interview. Table 4.1 lists the resulting feedback key points from residents and caregivers.

The setup took approximately two hours at the facility. During this time, the robot's functions were explained and a path was taught from the loading cable in the nurses' staff room to the common room, as shown in Figure 4.2. Although all doors are accessible and at least 1 meter wide, Pepper had difficulty maneuvering between rooms due to inaccurate positioning. While teaching the robot, instructions were displayed on the touchscreen. Especially in path-teaching mode, the robot had no artificial feedback, such as a sound or flashing LEDs, to signal that it had completed its task. Since the robot was moving away from the user most of the time, this was inconvenient. Therefore, the staff suggested the possibility of a split-screen application in future systems, showing the instructions on a handheld device. Automatic movement to a loading station and additional automatic loading was requested.

In general, the residents liked Pepper. It had an uplifting effect on the residents. The reason for this, besides the entertainment functions, was that the residents could tell the robot everything without being laughed at or gossiped about with others. In the beginning, one person was against the system, but as the days went by, the acceptance improved. One reason for this was the observation of the system over several days, to see that the robot stops in front of other people when it moves and does not hurt anyone.

The usability has been criticized in some points. The robot is not usable in large groups where multiple users want to interact with the system, as it is designed to interact with one person at a time. The test environment was naturally noisy and busy which is why the robot was often unable to recognize commands. This issue is signaled by Pepper when it turns its head from one person to the next. At that moment, the attention switches to the person whose face the recognition algorithm recognizes best. In some applications, the robot unintentionally switched to English because no German version of the function was provided. Most of the residents do not speak English. Many wanted to try the handshake function. The robot has touch sensors on the back of its hands, but they were not used in the handshake function. After an internal countdown, it began shaking hands, which caused irritation. When using the Memory app, Pepper kept moving the upper body where the touchscreen is mounted. This made it difficult to select cards. Furthermore, residents wished Pepper had more conversational capabilities. With the ongoing development in smart devices technology, residents and caregivers are accustomed to newer technology. In particular, the robot's responsivness was outdated due to their daily use of

voice assistants such as Alexa (Amazon), Siri (Apple), or Google Assistant (Google). Finally, residents mentioned that Pepper is naked. They wanted to dress him up for fun. The customization of parts could increase the acceptance of the humanoid robots, and is provided by additional manufacturers for Pepper [148].

In terms of accessibility, hearing-impaired people were unable to interact with the robot because it was too quiet. In addition, the robot needs to be build more stable because it could be kicked unintentionally since some disabilities, such as spasticity, do not allow the constant control of body movement. Another issue is coverage by private and government health care systems. The cost of a robot that exceeds a certain price cannot be covered by the institutions. This is also true if the system is to be used by individuals living on their own.

Finally, feedback was sought on specific tasks that the robot should help with after trying the robot for eleven days. Household tasks should not be done by the robot, since in this type of all-day care, the residents are responsible for themselves. Instead, the robot should perform pick and place tasks. Implementing Bluetooth and smartphone applications could help, as one of the most common requests was to play music. This would also allow bedridden residents to call the robot. One suggestion to improve the robot was that it should be able to push a wheelchair, making it easier for wheelchair dependent residents to get around.

4.3.2 Assisted Work

The workshops of WfB Haslach gGmbH are located in Elzach, Hausach and Steinach in Baden-Württemberg, Germany. These facilities produce various products such as printed pens and shirts, wallets, key chains, sewing kits and parts for automotive and other industries. The work can be described as predominantly assembly work. Three main differences from general production lines have been identified.

- **Plant utilization:** Utilization capacities fluctuate due to varying order situations caused by delivery bottlenecks and peak production times specified by the customer.
- **Daily quantities:** Due to the disabilities of the workers, the quantity may vary from day to day. Unlike most industries and due to the low level of automation, absent workers cannot be replaced by other workers. Most workers have a limited range of varied work in which they specialize.
- **Various production lines:** The aim of this company is to promote the employment of disabled people. Therefore, most of the work is done manually. Many machines were presented to assist the workers in production, such as sewing machines, printing machines (screen and pad printing), and devices for aligning parts and assisting with assembly tasks.

The role of the workers varies. most of manual work is done by people with disabilities. Trained supervisors monitor the work process, assist workers in the process, but also in activities such as assisting with toileting, and perform tasks such as quality control. Due to the lack of digitalization and the limited use of high-tech control and measuring equipment, quality control is also done manually. As a result, this task has to be interrupted when another person needs assistance, resulting in a restricted workflow. This is one of the company's biggest challenges. Based on these statements, the requirements for the robotic system were identified and presented in Table 4.2.

The second step was to develop and evaluate solutions. Using a state-of-the-art camera, solutions were provided for quality control of various products. For example, a computer vision system was set up to verify that a print on a pen was applied correctly by color and position. The algorithms used for measurement were successful. Whether the systems have the desired effect is part of future work.

Table 4.2: Functional and non-functional requirements for robot application in assisted work.

No.	Requirement	Description	Funct.	Non-Funct.
1	Collaborative	The robot must stop in the event of unintentional contact with the user [149].	x	
2	Unsupervised use	The robot must perform automated tasks.	x	
3	Graphical or manual teach mode	Non-expert users must be able to reprogram the robot for new tasks without extensive training.	x	
4	Implementation of Quality control	Cameras and sensor systems must be deployable.	x	
5	Mobile setup	The robot must be movable to be used at different work stations.	x	
6	Automatic calibration	When the robot is moved to a new task/position, it should be able to recalibrate its base and teach the new position by itself.	x	
7	Adaptability in speed	The robot must be adaptable to the speed of the user. The robot should wait in certain positions for the user to take the parts.	x	
8	Non-intimidating appearance	Users should not feel threatened by the robot.		x
9	Robotic Arm	Unlike other types of robots, the robotic arm provides the necessary flexibility to be used in a variety of environments.	x	
10	Robust against high external forces	Rough contact with users cannot be excluded. Therefore, higher external forces must be expected.	x	
11	Robust against lighting changes	Quality control cameras need to perform equally well in a variety of lighting conditions due to the mobile nature of the setup.	x	

4.4 Discussion

4.4.1 Comparative Overview

Pepper was released in 2014, which could be a reason for some of the negative key points such as its outdated language model. Talking with the robot about personal problems was also discussed in other studies [150]. Due to the continuous progress in speech recognition, new applications have been developed for Pepper, presenting an interface for using ChatGPT [151]. In the work of Sheridan [152] and Papadopoulos et al. [153], it was found that robots gave ore unsatisfactory answers when asked simple questions. For more complex questions, the robot's performance was similar to that of a human. As noted in [152], collision avoidance receives special attention. This requires not only localization, but also that the robot is aware of its environment. As seen with Pepper, collision avoidance worked well, but localization of the robot in the building had shortcomings. In recent years, Simultaneous Localization and Mapping (SLAM) algorithms have greatly improved through the use of more accurate sensors and more robust algorithms, providing suitable solutions to this problem.

Interaction is affected by further parameters. For example, gestures and reactions are crucial for human-robot interaction [152]. Improving the touch sensors in the hands could not only improve interaction scenarios like shaking hands, but also provide a basis for giving the robot information needed for grasping tasks. Using Pepper as an example, the grasping strength must be enhanced in order to grasp objects. The necessary trajectory planning along with velocity profiles, orientation, and prediction of the target location influence user experience and acceptance of the system [152]. In addition, it has been shown that physical interaction with a soft and warm robot is preferred [152]. In visual design, perception depends on the individual [153]. Insights gained by studying the "uncanny valley" effect showed that a mixture of abstract and human-like appearance can help. As an enabler for social humanoid robots, personalization by adapting to the user's tastes and preferences has been mentioned as it is available for Pepper [153].

Additional features such as walking and supervisory tasks were suggested by the caregivers and residents. As the question arose whether a robot would be a suitable walking aid, Karunarathne et al. showed in a study that elderly people would prefer a humanoid robot over walking alone [154]. However, the humanoid robot used lacks a suitable design as a physical aid for people with motor impairments and its inability to move on roads or even on gravel. The question is whether the user would accept the robot over a human companion, and whether cognitively impaired people would react in the same way as the participants in this study. In general, a robot used as a walking aid should be small, equipped with handles, and preferably have a seat and wheels capable of traversing different floor surfaces.

In the workshop environment, the requirements analysis showed differences from general industrial production, which makes the ubiquitous use of robots more important. The focus is on maintaining manual jobs to ensure prospects for people with disabilities. In particular, the robot must be used at different workstations, assisting people without fixed cycle times, but also adapting to the abilities of each user, resulting in a high level of social behavior. Since the people living in the community housing also work in the company, the crossover effects between social and collaborative robots in the user experience are of interest. No relevant literature was found regarding this topic. It is assumed that there should be no significant drawbacks when using both systems in one facility, since the robotic arm used in the assembly tasks is very different from humanoid robots such as Pepper used in this study (Figure 4.3). A challenge in using both robots could be that Pepper actively avoids unintentional contact with the user and its grasp strength is low. On the other hand, industrial and collaborative robots can grasp more strongly and move faster. The user may have difficulty estimating the robot's strength. This could result in the person refusing to cooperate with the robot.

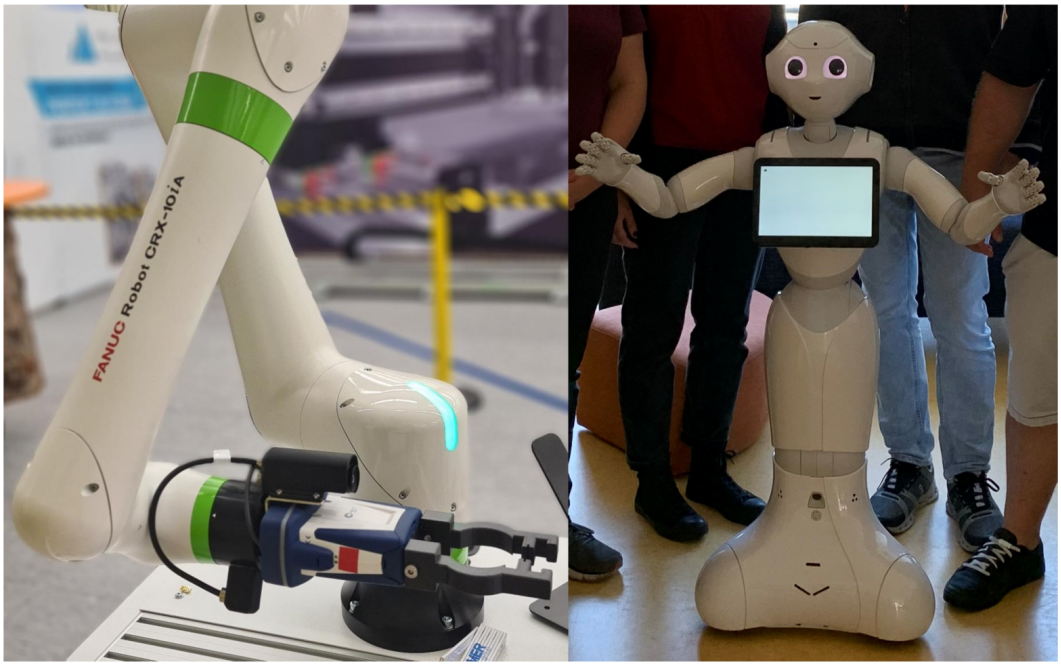


Figure 4.3: Use of robots in different environments. Left: Example of a collaborative robot with a camera for assistive work, Right: Social robot Pepper used in assisted living.

4.4.2 Ethical Questions

Working with people with cognitive impairments, the institution mentioned that using the robot as a walking aid would improve the quality of life for some residents. In this case, the robot should respect the human's freedom of choice, but also ensure that the environment is safe and that the human does not get into risky situations. This raises the question of whether a robot can take over this care task from a human, and what safety functions need to be integrated to meet these requirements.

In the context of privacy, this work revealed a requirement that contrasts with ethical choices. Including applications that allow the robot to be called by a bed bound person would improve the interaction possibilities between the user and the robot. However, one requirement was that the robot should not enter private rooms. Exceptions need to be defined. In addition, data security is an ongoing issue in robotics due to the use of cameras or the recording and processing of human voice in private environments, especially regarding cloud-based solutions as mentioned above. Camera imagery and LiDAR data have a huge impact due to their detailed imaging of people and environments. This raises concerns in combination with recent legal uncertainties, the increasing autonomy of robots and the replacement of human interaction [155]. With more sophisticated and faster ways to process information, the definition of human-centered data changes. For exoskeletons or adaptive robots, muscle activity detected by the robot or user behavior can be included as part of the human-centered data [155]. To find the right balance between not providing needed functionality due to a potential lack of data and creating a risk to personal security, Fosch-Villaronga et al. suggest involving real users in the process and being transparent about the data collection process so that people can decide for themselves whether they want to share the data [155]. Not only the opinion of the disabled person and caregivers should be sought, but also that of relatives, as in some cases they have a care directive. Including such a number of stakeholders the use of private information should be as clear as possible [139]. The experience from this work shows that each participant (manager, caregiver, and resident) was happy to have this opportunity to help. Nevertheless, it has to be clear that there is a huge responsibility when working with e.g. dementia patients or participants with disabilities [153].

4.5 Conclusion

This paper presents a case study for social robots in assisted living using the example of Pepper, which was tested for eleven days by residents and caregivers in a community housing complex for assisted living, and for general robotic application and automation in assisted work. For both cases, design and usability requirements were extracted based on user experience. The results were analyzed in comparison with existing literature. The comparison showed similar results. Ethical issues were raised by the organization and the company, which were not found in the literature.

Acknowledgement

This version of the contribution has been accepted for publication,a after peer review (when applicable) but is not the Version of Record and does not reflect post-acceptance improvements, or any corrections. The Version of Record is available online at: https://doi.org/10.1007/978-3-031-61315-9_20. Use of this Accepted Version is subject to the publisher's Accepted Manuscript terms of use.

Chapter 5

Publication III: Eye-Tracking-Driven Shared Control for Robotic Arms: Wizard of Oz Studies to Assess Design Choices

Summary: Based on the findings of the scoping review in Chapter 3, a low workload and an easy-to-learn control system contribute to a positive user experience. In this chapter, an eye-tracking-driven control system was designed to meet these criteria and to enable the evaluation of usability and functionality. A shared control approach in which the robot has a certain level of autonomy was chosen to reduce the duration of the active control phase. This reduces fatigue and cognitive workload. The system employed a Wizard of Oz design to collect feedback from people with disabilities, their families, and their caregivers. A rapid prototyping approach was used to identify shortcomings through surveys and avoid implementation drawbacks. This approach was used to gain subjective perceptions of the system from stakeholders and to avoid the pitfall of excluding people with disabilities. A second study was conducted since this approach could not estimate functionality. In the study, non-disabled people physically interacted with the system to determine if accuracy issues would result in incorrect selections and evaluate the impact of natural gaze behavior.

The control involves fixating the gaze on an object for a certain amount of time to select the object and, thus, the task. These approaches are called dwell-time approaches. Bounding box overlap and small sections in object segmentation severely limited functionality, thereby reducing performance in the selection process. Accuracy errors exacerbated this issue, sometimes making it impossible to select an object when the participant gazed through the glasses near the periphery of the field of view. Comparing the video representation in the survey with real-world interaction and the perceptions of people with and without robot expertise offered new perspectives on the different perceptions of the system and confirmed the results found in the literature.

One object can indicate multiple task intentions, causing uncertainty in task selection. The goal was to increase the number of available tasks from five in the WoZ system to over 16. The tested control method poses a high risk of selecting the incorrect task at this scale. Accordingly, the control method presented in Chapters 6 and 7 was modified to improve robustness in task selection and enable scalability.

Contribution: Anke Fischer-Janzen was the primary author of this publication and was responsible for the conception, implementation, evaluation, and preparation of the manuscript. Thomas M. Wendt, Daniel Görlich, and Kristof Van Laerhoven contributed through supervision, methodological discussions, and manuscript review.

Eye-Tracking-Driven Shared Control for Robotic Arms: Wizard of Oz Studies to Assess Design Choices

Anke Fischer-Janzen
Offenburg University of Applied Sciences

Thomas M. Wendt
Offenburg University of Applied Sciences

Daniel Görlich
Offenburg University of Applied Sciences

Kristof Van Laerhoven
University of Siegen

Abstract

Eye-tracking-driven controls for assistive robotic arms provide people with severe physical disabilities with intuitive interaction opportunities. In this context, shared control can improve acceptance of the robot by partially automating task execution. Based on recent literature, we present a Wizard of Oz design for shared control driven by eye-tracking. This approach allows for the rapid exploration of user expectations to inform future design iterations. Two studies were conducted to assess user experience, identify design challenges, and determine ways to improve usability and accessibility. The first study involved people with severe motor disabilities (PSMD) and consisted of an online survey. The second study aimed to gain technical insights through direct robot interaction to provide a comprehensive overview of the findings.

Keywords: Assistive Robotics, Eye-Tracking, Robot Manipulation, Assistive Device, User Experience, Disability, Wizard of Oz

5.1 Introduction

People with severe motor disabilities (PSMD) often rely on hands-free control modalities, such as Brain-Computer Interfaces, sip-and-puff systems, and eye-tracking interfaces. These technologies have been successfully used by individuals with conditions such as locked-in syndrome, cerebral palsy, late-stage amyotrophic lateral sclerosis, or multiple sclerosis. As control inputs for assistive robotic arms, they can support independence, self-determination, and privacy in daily life, reducing reliance on family members and caregivers [12, 156, 157].

A shared control allows users to give the robot high-level commands, such as selecting an object or goal, by automating task sequences. In eye-tracking-driven approaches, advances in AI allow us to move beyond manual mappings toward intelligent interfaces that interpret gaze patterns to extract task intent [90, 11, 158]. Environmental context, such as the presence of objects, can further improve autonomous decision-making [22, 20]. These advances have led to the development of various robot control methods for Human-Robot Interaction (HRI) that differ in terms of control strategies [159, 12]. With a higher level of robot autonomy, challenges such as controlling three-dimensional robot movement with two-dimensional eye movements can be resolved. Bypassing the need for mode switches, as required for direct control, to cover the 3-dimensional robot space, or independent joint control, can reduce workload and decrease the concentration required for the task [12, 160, 161]. Shared control approaches have been shown to positively impact interaction perception and reduce task completion time [15, 160]. A challenge associated with a high level of automation is maintaining the sense of agency. Community feedback is essential to improve design and evaluate usability and accessibility [48, 157, 162], as well as to uncover influences and design controls with an appropriate level of autonomy. It provides firsthand insights into user expectations, thus enhancing acceptance rates [162, 163].

The novelty of this work is characterized by a detailed look at usability and functionality regarding robot acceptance parameters in respect to articulated robots used to assist in daily life. Current task

intent recognition methods are based on object labeling done by object detection or segmentation, which we evaluated in this context on feasibility. This work incorporates community feedback on an initial shared control design to identify challenges and guide improvements. A Wizard of Oz design of the eye-tracking-driven control was developed and used in two studies. The first study was an online survey distributed to PSMD, their families, friends, and medical professionals. The goal of the study was to assess everyday usability and identify task preferences. The stakeholder group was defined as PSMD who experience significant motor impairments in their arms and torso. Previous research has shown differences in the perception of virtual versus real robot interactions [4]. A second study was conducted, in which participants without motor impairments and with varying levels of robotics experience interacted with the physical system. This study focused on deriving technical characteristics that can inform the constraints of a future control algorithm, such as feasible dwell times and accuracy requirements. The main contributions of this work are as follows:

1. Presentation of an initial eye-tracking driven shared control design leveraging object detection algorithms for daily tasks.
2. A functionality and usability evaluation, incorporating feedback from potential users, their families, medical professionals, and non-disabled participants with and without prior robot experience.
3. Refinements and task preferences for the initial shared control design, derived from insights obtained in both studies.

5.2 Background

5.2.1 Gaze as control input in shared controls

"Shared control in assistive robotics blends human autonomy with computer assistance, thus simplifying complex tasks for individuals with physical impairments" [164]. Such a system can achieve lower performance times and error rates when the user provides strategic intent and the autonomy supplies continuous, low-level stabilization, obstacle avoidance, and motion refinement [16]. Advances in sensing, inference, modeling, and learning methods have led to novel developments in this research area [165].

Gaze can be used as an implicit or explicit control input for shared controls [8]. As an implicit control input, natural gaze patterns are used to recognize or predict user intent. For example, gaze behaviors such as fixating on task-relevant objects before naming or manipulating them can be used and predicted before the desired task even begins [166, 92, 11, 158, 165]. Huang and Mutlu demonstrated an example of such controls, in which ingredients were predicted by gaze patterns [65]. Admoni et al. created a shared control system based on a POMDP (Partially Observable Markov Decision Process), which predicts single tasks from gaze patterns in an unstructured environment [13]. Zhang et al. presented another approach based on the usage of a large language model (LLM) and vision foundation model (VFM), that recognizes user intent by interpreting labels created by the VFM and using the LLM to find common object combinations [22]. Wang used machine learning to recognize actions based on 3D gaze-related features to control a robot [167]. Lastly, Farhadi et al. presented a review of BCI-related controls and shared control approaches [16]. Challenges in predicting patterns of natural gaze can arise from multiple cognitive processes occurring in our brain at any time. As Belardinelli states, this issue worsens when performing activities more complex than grasping or selecting an object [11]. If incorrect or imprecise predictions are made, collisions or large penalties in completion time and user preferences can occur [89]. As Cao et al. point out, timely and accurate estimation of intentions and, consequently, prediction remain major challenges [10].

Gaze, as an explicit input, can vary in its function in control. The first eye-tracking controllers for robotic arms were developed as early as 2001 [39]. Since then, several systems have been developed, but only a few have become commercially available [12, 168, 169]. One example of explicit control input is directional control, in which the robot moves in the direction of gaze [70, 51]. Graphical User Interfaces (GUIs) are available in which the robot is moved with gaze-selected buttons, taking advantage of more accurate stationary eye-tracking systems [52, 48]. Such controls are easy to learn, provide a good sense of agency, and reduce cognitive load [52, 59]. However, controlling each of the robot movements separately is more time consuming. A popular approach, as outlined in Belardinelli's work, is to use relevant objects in the scene and define them as Areas-Of-Interest (AOI). Fixations within these areas can be evaluated and, together with a certain dwell time, used as a selection method [18, 45, 17]. We refer to the work of Fischer-Janzen et al. and Suryadarma et al. for examples in realization, due to the number of related works [12, 168]. A major limitation of these works is that they do not allow for the subsequent selection of multiple tasks, which is a major requirement for daily use, as demonstrated by Hagengruber et al. [5]. In this work, gaze is used as explicit input by focusing on the desired object.

Challenges can occur in both implicit and explicit control inputs that need to be considered:

- **Unintended selection:** When the gaze is used as a pointing device, e.g., to select an object, rules must be applied to ensure that not all fixations do not trigger selections. These unintentional selections are described by the Midas Touch problem [170, 46, 48]. Reasons for this problem include unintentional and reflexive eye movements, overlapping objects, distraction, stress, and the limited accuracy of the eye-tracking device [31]. When objects are close together and the eye tracker's accuracy is insufficient, the gaze may be tracked to the wrong object [46].

- **Context misinterpretation:** The aforementioned parallel cognitive processes can result in task ambiguity. In an interview conducted by Herr et al. [171], participants stated: "Well, now I look at you [the interviewer], but I actually focus on the bottle behind you. That would be difficult for the robot to recognize." This behavior suggests that attention to a particular location is sometimes unrelated to intention or cognitive process.

- **Physiological factors:** There are some physiological factors that affect the accuracy of the gaze. Neurological conditions, fatigue, drooping eyelids, and strabismus can affect the functionality of eye-tracking devices [31].

Such challenges can lead to frustration and a high cognitive load for users trying to avoid them [170, 46].

5.2.2 Characteristics in gaze-based shared controls

Our system design was inspired by the impact factors on user acceptance of shared control found in the current literature. Many characteristics and technology acceptance models exist [172]. Due to the multifaceted nature of user acceptance, we present only the factors that influenced our design. For more information on characteristics influencing user acceptance, see the works of Smarr et al., Rahman et al., Riek et al., and Beer et al. [173, 174, 175, 172]. We introduce three categories to facilitate an overview of the results: Technical challenges, level of autonomy, and user preferences.

Technical challenges include ensuring the system's reliability and functionality. The system must be robust in changing or unorganized environments, especially when the robot is mounted on an electric wheelchair [165, 176]. The system's capabilities related to task performance also define its functionality in everyday life. We discussed such challenges in Section 5.2.1. The subtasks necessary to correctly complete a task must be understood and set in the proper context relative to each other.

The *level of autonomy* is a subject of ongoing debate and can range from tele-operated systems to fully autonomous robots [177]. Recent studies have shown that more autonomy does not necessarily align with user preferences [6, 15]. If a robot with a high level of autonomy produces more errors in task execution, users tend to reject the system [15]. This can lead to a loss of trust or a reduced sense of agency due to the system's unpredictable behavior [6, 178]. Losey et al. summarized this as a duality: More automation reduces the user's burden when performing tasks, but it also limits the mutual support when it ignores the user's intent. The autonomous behavior reduces the operator's workload when performing repetitive tasks [165]. As Loke et al. state, high mental workload contributes to the abandonment of assistive devices [8]. Due to the potential shift in cognitive resources toward the task and away from precise robot movements, control can benefit from being easy to understand [179, 165]. However, this may vary depending on the complexity of the task [176, 97, 164], leading back to the initial question of what level of autonomy is most efficient and reliable in assisting the user.

Users' preferences vary based on personal metrics such as the user's age, level of disability, living arrangements, and familiarity with technology [176, 173, 6]. Incorporating user preferences influences perception of and satisfaction with the system [180, 163]. Parameters such as speed can influence the user perception [15]. Another example is the participant's preference for handing over an object with a particular orientation [180]. The ability to adapt the system seamlessly to users' preferences regarding the level of assistance has given rise to shared autonomy approaches. Those are based on the autonomy's understanding of the user's intentions and of the surrounding environment [165]. In this context, research focuses on arbitration strategies that adapt the level of autonomy to the user's preferences, as demonstrated in the works of Dragan et al., Losey et al., and Jain et al. [181, 89, 182]. The arbitration scheme, together with the interaction paradigm, describes the collaboration between human and robot [16].

When designing well-balanced shared control, a trade-off must be found for all technical and ethical factors. Section 5.3.1 presents a solution that considers and balances the different influences.

5.2.3 Tasks in assistive robot manipulation

Edemekong et al. describe Activities of Daily Living [(ADLs)] as basic routine tasks that most healthy individuals can perform without assistance [183]. ADLs are often used to evaluate a person's level of disability, as in the WHODAS 2.0 [184]. Nanavati et al. used ADLs in their review to compare the responses of individuals requiring ADL assistance with the work of researchers in the field of assistive robotics. A major misalignment was found. While most researchers focus on mobility and eating, people in need of assistance state that they need help with going out, managing money, doing housework, preparing meals, bathing, and dressing [6]. Petrich et al. took a different approach, examining the behavior of people in everyday life and uncovering results quite different from those previously mentioned. In their study, tasks such as reaching to pick up and place items, opening and closing doors, and using switches and buttons were most important, according to video analysis. These results differed from the participants' survey answers [4]. These results highlight that different approaches can reveal various aspects of task prioritization and preferences in HRI. Smarr et al. explored robot acceptance among the elderly and found that the elderly accept robots more readily when they assist with learning new skills and hobbies. However, human support was preferred for communication, which emphasizes the importance of social contact [173]. The examples presented above show that assistive robotic arms cannot directly assist with activities such as bathing or managing finances. It is important to distinguish between robotic arm-related tasks and others.

As stated by Hagengruber et al., the purpose of assistive robotics is to carry out various tasks in succession in order to depict a realistic daily routine [5]. They presented an sEMG-based control system that could perform a drinking task, including retrieving a cup from a drawer and pouring a drink. Chi et al. demonstrated that dividing complex tasks into subtasks facilitates accomplishment.

For example, they divided an eating task into subtasks such as getting a spoon, scooping food, and bringing it near the user's mouth [185]. Using gaze as a control input, one can obtain gaze patterns specific to each subtask that can be used to train AI models [186].

Robots can perform daily tasks, such as household chores and personal hygiene, as well as medical procedures [171, 174, 26, 174, 187]. The preferred tasks vary depending on different influences, such as the user's attitude towards a task. Delegating household tasks to robots is one example of this [171]. Similarly, Petrich et al. suggested automating tasks that are not prioritized by the community to reduce cognitive load [4].

Lastly, one must determine whether a robotic arm is the right tool for the task at hand. For example, Petrich et al. stated that robots can help with using electronics, such as smartphones; however, it is more convenient to use a hands-free phone [4]. Similar results were found for toileting, where people preferred the aid of caregivers and smart home applications. Hagengruber et al.'s work, in particular, demonstrates the advantage of using multiple assistive devices together. For example, it simplified opening a drawer by holding the handle with the robot and moving the wheelchair backward [5].

In addition to the work of Nanavati et al. [6], our review revealed that the four most frequently mentioned tasks, suitable for assistive robots, were food related tasks (serving a meal, eating) [26, 176, 188, 189, 163, 15, 162], drinking related tasks [26, 176, 188, 189], picking up objects [26, 176, 188, 189, 6], and personal hygiene [176, 188, 6]. Personal hygiene included various tasks such as shaving [26, 189], washing the face and hands, brushing teeth, combing hair, shaving, applying makeup [188], and dressing [163, 6]. The importance of being able to manipulate a wide variety of objects benefits the ability to get around, e.g., with an electric wheelchair, by being able to move blocking objects out of the way [6].

Other miscellaneous tasks were mentioned with taking medicine [6], working [6], leisure activities [176, 162], turning on and off switches [26, 189], opening and closing doors [26, 169, 189], scratching oneself [26, 189], and making tea. Sources from 2003 show that changing CDs and removing paper from fax machines are examples of how task preferences change over time [26, 189]. As the accessibility of smart home applications continues to evolve, it is expected that more tasks will become less important for robot assistance, such as flipping light switches or opening doors.

Based on these results, tasks were selected for the Wizard of Oz design. In Section 5.4.1, the literature findings were used to group the responses from the online survey. In Section 5.5.1, they were used to compare the results in terms of changes in task prioritization over time.

5.2.4 Participation in assistive robot design

The previous section presented characteristics that influence user acceptance and robot abandonment. Several authors argue that involving the community and people with disabilities in the design of assistive devices is more advantageous than involving people without disabilities. Including PSMD in the design process of assistive devices helps designers gain insight into the lives and needs [157, 176, 162]. Arboleda et al. point out that end users' perspectives can differ from those of other groups, such as caregivers [190]. As Stalljann et al. demonstrate, control inputs are experienced differently by disabled and non-disabled participants [48]. Most studies on eye-tracking controls recruit non-disabled people as participants [12]. Reasons PSMD are rarely included in studies include recruitment challenges, small sample sizes, and transportation logistics [162].

Nanavati et al. [162] discuss approaches based on three-dimensional key features to determine suitable study designs for achieving the study's objective. The first dimension determines whether participation is at the individual or community level. Individual participation provides the opportunity for deeper insight into the user's life and preferences, while community work provides insight from many participants. This dimension also affects the time required to conduct the study and the connection between participants and experimenters. The second dimension highlights the logistical effort required

of participants and researchers. PSMD participants may require assistance with transportation to study sites, which can facilitate conducting studies at the participants' locations. Arboleda et al. stated that conducting the study itself might disrupt participants' daily routines, which could induce stress [190]. Therefore, other possibilities can be considered, such as online interviews. Depending on the robot setup, it may be easier to conduct studies at the participant's location to reduce the need for special transportation and availability of medical professionals. The third dimension describes the benefits to researchers and the community. PSMD, their families, and medical professionals are intrinsically motivated to participate in the design process of assistive devices and robots. They can benefit and contribute by providing insight to adapt the systems to their needs.

These findings demonstrate the value of involving PSMD in the design process and influenced the study design in Section 5.3.3.

5.3 Methods

In Sections 5.1 and 5.2, we presented the influencing factors and user challenges that affect acceptance of eye-tracking-driven shared controls. These factors and challenges were incorporated into the system design presented in Section 5.3.1. Several groups of interest were interviewed with two study designs, to obtain broader feedback. The setup described in Section 5.3.2 was used in both studies. An online survey was sent to stakeholder groups, their families and medical professionals. The results are presented in Section 5.3.3. Measurements were conducted to verify technical functionality and participants' perception of the system interacting directly with the robot (see Section 5.3.4).

5.3.1 Shared control design choices

The initial idea for the system is to include a shared control, which automates parts of the robot movement, such as trajectory planning, grasping, and object manipulation. The level of autonomy is presented according to the guidelines of Beer et al. (2014) [177]. The following considerations were used to create the Wizard of Oz design and describe the interaction to the participants.

What task is the robot to perform? This shared control system is designed to assist PSMD with daily tasks. Ongoing work is developing a framework that can be expanded to include new ADL, thereby increasing the range of tasks it can perform. The first approach includes the tasks of picking up food with a fork, turning on a light switch, picking up and placing a block, scooping from a bowl, pouring water into a glass, and handing it to the participant. These tasks have to be performable in unstructured environments, such as at home and in public places like restaurants.

As shown in Section 5.2.3, these tasks are preferred for robotic assistance in everyday life.

What aspects of the task should the robot perform? The chosen input modality is eye-tracking, because eye movement is often unrestricted or only slightly restricted in cases of locked-in syndrome or cerebral palsy, in contrast to other motor functions, such as arm movement. The robot is controlled by fixating the desired object, and the robot decides which available task to perform based on the selected object. Selecting multiple objects with the gaze makes task selection more robust. For instance, looking at a cup could indicate that the user wants to drink from it, move it, fill it, or combine available tasks. Adding the ability to combine the cup with other objects, such as a bottle or a location, distinguishes the task as refilling the cup or placing it in the selected location. Based on the information about the object and the task, the robot performs the task autonomously. The robot determines the object's location, selects a suitable grasp related to the task, and calculates the optimal trajectory.

To what extent can the robot perform those aspects of the task? The user is always in control of the robot in order to maintain a sense of agency (discussed in [6]). Validation methods are planned to intervene in case of incorrect task execution before the robot begins (suggested in [178]). Settings are planned to reduce the robot's velocity, if desired (suggested by [15]). Implementing dwell time for

selecting objects has the potential to reduce incorrect interactions with objects through natural gaze behavior, thus minimizing the Midas Touch problem.

What level can the robot's autonomy be categorized? According to the classification of Beer et al. [177], robot autonomy is categorized as Shared Control with Human Initiative.

How might autonomy influence HRI variables? The system must be reliable and functional. User safety is part of ongoing discussions and must be ensured. The system is designed to be easy to learn and use. In real-world settings, objects are selected by focusing on them. However, the accuracy of the eye tracker limits the correct selection of the desired object when gaze is tracked outside the area of interest (AoI). This raises the question of how accuracy errors influence the functionality of this approach. This will be tested using the Wizard of Oz setup.

5.3.2 Wizard of Oz design

Table 5.1: Wizard of Oz presentation after Riek [191] stating experimental components to the robot, user, wizard and to general information.

Experimental Component	Online survey	Hands-on study
Robot	7-DoF Kinova Gen3	
Level of autonomy	Teleoperation	
Robot capabilities	Food item, scooping food, filling cup	Food item, pick and place, filling cup, light switch
No. of participants	34	24
Demographics	Age: 40 ± 27; 14 PSMD, 9 family members, 11 medical professionals; 14 male, 17 female, 3 diverse	Age: 30.5 ± 8; 11 non-prior experience group, 13 prior-experience group; 7 female,17 male
Instructions	See Section 5.3.3	See Section 5.3.4
Convincing simulation	Not asked	Yes
Wizard	Experimenter took the role of the wizard	
Familiar with hypothesis?	Yes	
Production variables	-	See Section 5.3.4
Recognition variables	-	See Section 5.3.4
Wizard error	Video representation, no variation between tests	Experiments were repeated when robot malfunctioned, occurrence: 1.
Location	Unknown, not asked due to privacy regulations	At the laboratory
Environmental constants	Unknown	Setup as shown in Figure 5.2
Scenario	As shown in video (see Supplementary Materials)	See additional information to scenario.
Iterative design process	No	No
Limitations	See Section 5.6	

An early system evaluation was conducted to determine whether the design meets the community's expectations for the found influencing factors. This approach shortens development time by intervening early in the event of low acceptance or functional challenges. In the Wizard of Oz design, the experimenter performs the autonomously controlled robot tasks, as seen in Figure 5.1. Table 5.1

presents the Wizard of Oz system according to the guidelines presented by Riek [191]. Figure 5.2 shows the five selectable tasks with the Wizard of Oz system. A video of task performance was created and implemented in the online survey to demonstrate the system's functionality (Section 5.3.3). This video is included in the Supplementary Materials. The system was also used in the hands-on study, where participants operated the system in a controlled environment (Section 5.3.4). Both studies were reviewed and approved by the ethics committee of the first author's institution.

Reason for omitting the item "Experimental Component" from the Wizard presentation of Riek et al.: The hypothesis for the robot was to reproduce each participant's trajectory and create as little variation as possible. Thus, the robot's software computed the trajectory planning. The points between the trajectories were programmed so that the overall task trajectory would be reproducible between participants. As Arboleda et al. indicated, robot behavior is perceived differently in virtual and real presentations [190]. Therefore, a crucial aspect of the behavioral hypothesis was to compare the two presentations. Distinct reactions of the participants were not previously hypothesized, since the goal of this work was to gain insights into user expectations and behavior.

Additional information on the item "Scenario": The objects were placed in predefined positions within the robot's workspace. All of the trajectories necessary for the automated task execution were hard-coded and could be selected by the experimenter by clicking on the program. The participant's safety was ensured by separating their workspace from the robot's.

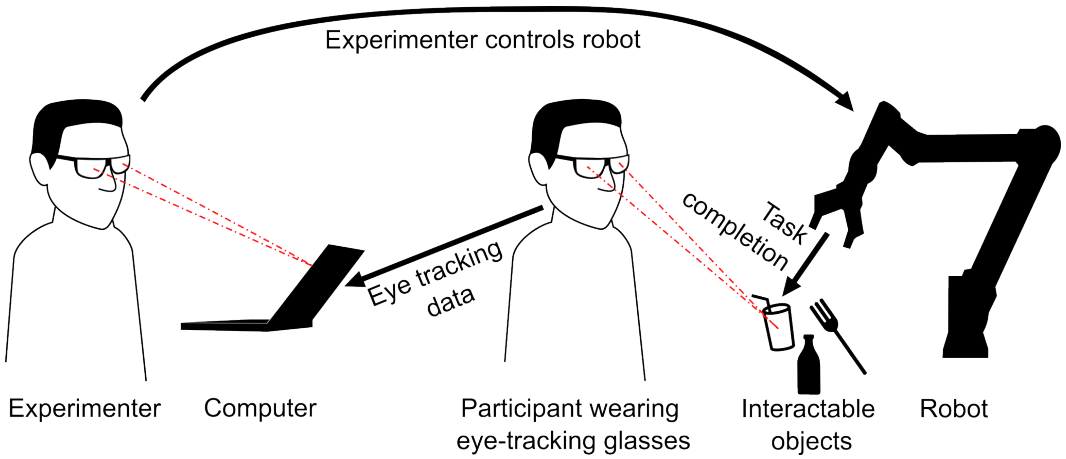


Figure 5.1: Wizard of Oz design: The participant selects a task by focusing his or her gaze on the desired object. The gaze is visualized on the computer by displaying the live-video of the eye-tracker. The experimenter starts the task via the robot's web interface. The robot then performs the task and returns to the home position to wait for further commands. The home position is used as indication that the participant can select the next task.

5.3.3 Online survey

The online survey was chosen as the data collection tool because it provided a standardized way to present information and collect feedback. The goal was to obtain comprehensive feedback from PSMD, their family members, and medical professionals in order to evaluate the system's usability and accessibility. Providing an overview of the system to the stakeholder group via online survey and video minimized the logistical challenges described in Section 5.2.4. This allowed immobile individuals to participate using smart devices. Unlike a real-world demonstration, where the authors could only visit a few facilities, people living at home could participate. Additionally, the survey could be completed at the participant's preferred time, which was convenient for those with complex schedules or who required assistance. Due to the anticipated limited availability of accessible computer controls, an

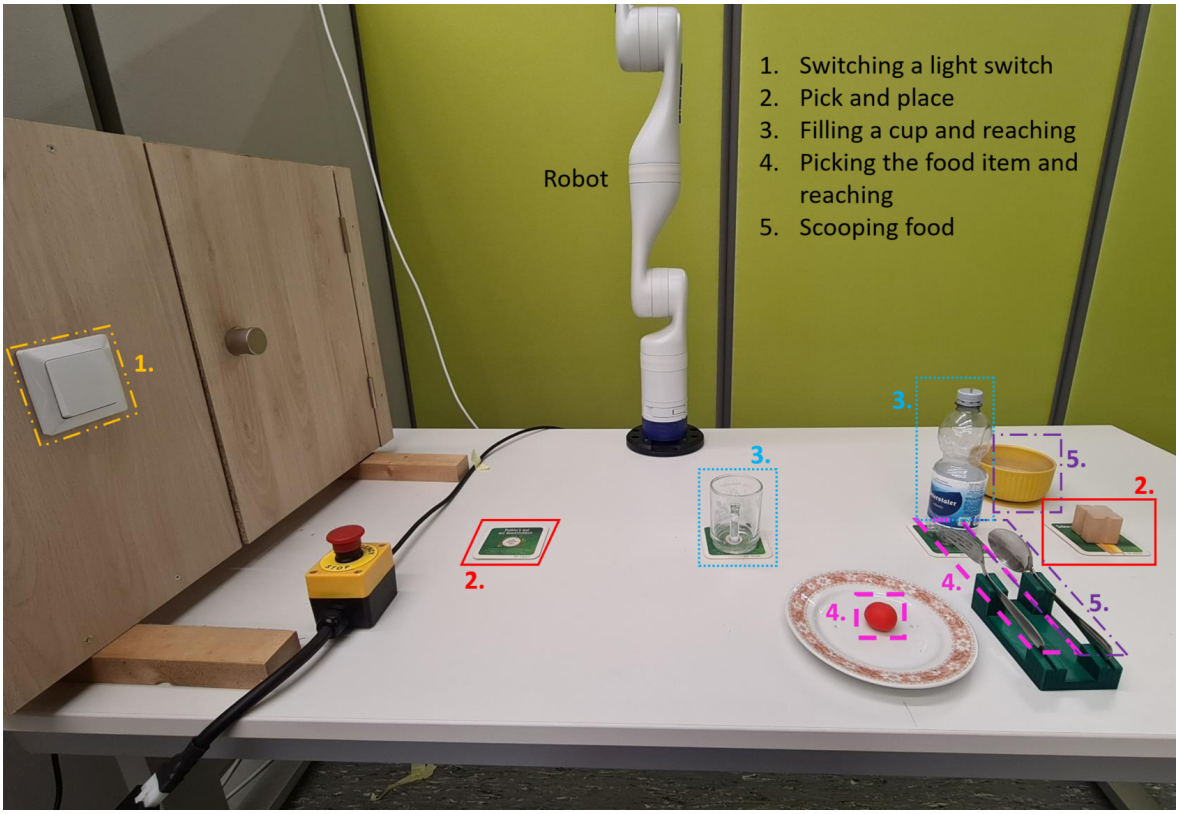


Figure 5.2: Nine interactable objects were presented with which five tasks were performed. Participants were shielded from visual distractions by green walls surrounding the experimental area.

extension was designed for family members and medical professionals to obtain more comprehensive feedback. The purpose of the online survey was to answer the following questions:

- Are potential users able to use the system?
- Which tasks should be prioritized to maximize the robot's usefulness in everyday life?
- Were the participants satisfied with the presented system?

Survey development and structure

The survey focused on investigating five topics, which are listed in Table 5.2. First, questions were asked about the participants' demographics to provide comparable and reusable baseline data. After the five topics were covered, participants were given the opportunity to provide feedback on the presented system and make general remarks. As stated in Section 5.2.4, perceptions of assistive devices can differ between end users and individuals in close contact with them, such as medical professionals and family members. Our goal was to receive a broad perspective on the prototype, which is why the survey varied between these three groups. Questions about the support situation were designed to evaluate the survey's accessibility. Poor accessibility can be indicated by how Q4(PSMD) is answered. A higher number of relatives and medical professionals who did not involve the person they care for can indicate time constraints. Furthermore, if everyone in Q4(relatives/friends) had stated that the person they care for had answered the survey, the answers would need to be weighted due to potential bias.

The topics of disabilities and assistive devices focus on whether the respondent is the intended audience. We implemented questions to reduce responses from people who misidentified their group

affiliation. The topics also focus on whether the respondent is familiar with assistive devices, including robotic arms. The literature contains cases in which users reacted differently to devices when they had prior knowledge. This may affect their reaction to the presented system. The questions regarding the robot presentation were developed to identify signs of abandonment. Robotic arms are not widespread. Upon inquiry, two companies that sell assistive robotic arms approved as medical devices reported selling around 1,200 models in Germany, Switzerland, and Austria. Since this strongly affects the system's perception, participants were asked if they had experience with robotic arms.

Lastly, task preferences were requested. As technology continues to evolve, the aim was to update the ongoing literature. As presented in Section 5.2.3, task preferences change quickly with the advent of new technology, such as smart home applications and hands-free smartphones. Some tasks, like using a fax machine, are no longer relevant. We asked participants to rank the importance of three tasks against each other.

Throughout the survey, we chose open-ended questions to receive rationale for the participants' answers to yes/no questions, such as reasoning for being unable to use the system. After initial development, the survey was discussed and refined.

Table 5.2: Breakdown of topics asked in the survey and related questions.

Question topic	PSMD	Rel./Fr.	Med. prof.
Support situation	Q4+Q5	Q4+Q5	Q4
Disability & assistive devices	Q6-Q8	Q6+Q7	-
Robotic arm use	Q9-Q12	Q9+Q10	Q9+Q10
WoZ satisfaction	Q13+Q14	Q11+Q12+Q14+Q15	Q11+Q12+Q14+Q15
Task preferences	Q15	Q13	Q5+Q13

As a result, the online survey consisted of 31 questions, divided into three groups, as well as a video describing the eye-tracking control of the robotic arm, structured as shown in Figure 5.3. In addition to the tasks, the video showed the calibration process. The exact questions asked are provided in Table 5.5 in Appendix 5.A. All survey questions, along with the participants' responses, are available in the Supplementary Materials. The survey was designed using LimeSurvey.

Distribution method

The survey was available online from September 18, 2023, to January 22, 2024. 32 organizations in Germany, Switzerland, and Austria were emailed to reach the stakeholder groups. Seven organizations, including FGQ (Paraplegic Support Association), LIS e.V. (Locked-In Syndrome Association), Diakonie (social services), Lebenshilfe (nonprofit association), and the Swiss Paraplegic Centre, shared the survey. The authors' private accounts and their institutions shared the survey a total of three times on social media. FGQ shared it in their private Facebook group, and the Swiss Paraplegic Centre posted about it in their private forum.

Participant instructions

First, participants were informed about the content and purpose of the study. Then, they were asked to give their consent to participate. Participants were then guided through the survey. Participants could skip questions at any time if they felt uncomfortable answering them. The video played when clicked on, and could be repeated if necessary. To offer different resolutions, participants could also open the video on YouTube.

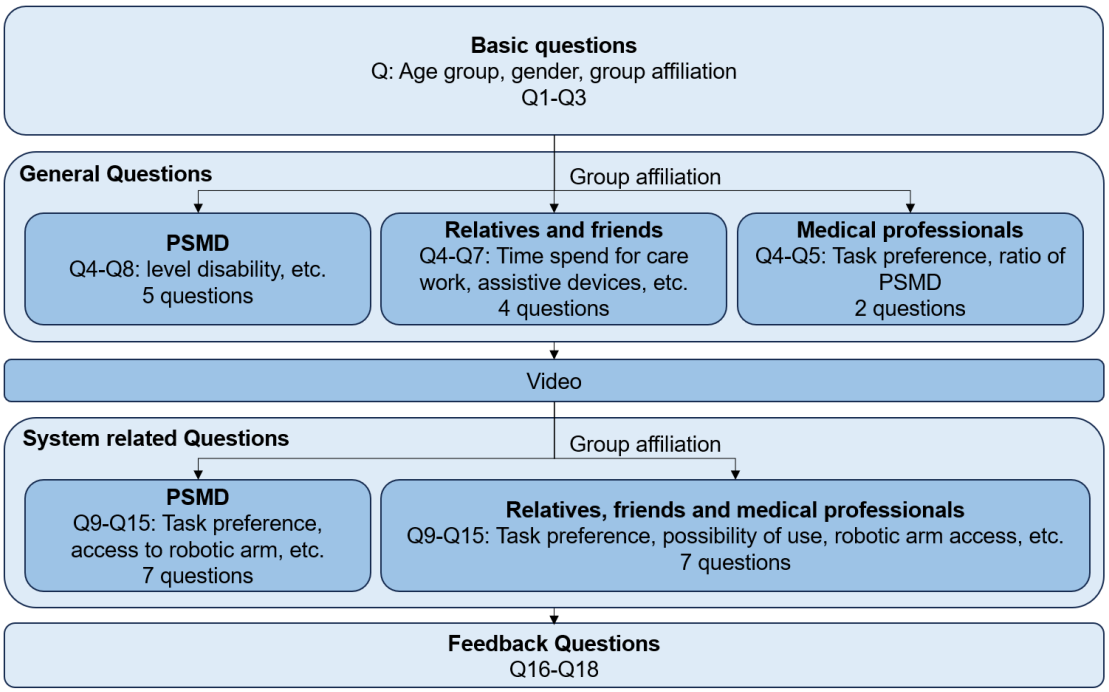


Figure 5.3: Distribution of questions asked for each group. Q outlines the context of the questions and the number of the questions for each category is given.

Evaluation method

A frequency analysis was performed to evaluate users' perceptions of the system and their desired tasks. As introduced in Section 5.2.3, the tasks were categorized as follows: eating, drinking, picking up objects, environmental interactions, personal hygiene, and miscellaneous.

5.3.4 Hands-on user study

In addition to the stakeholder evaluation of the system, a Wizard of Oz design was employed to analyze real-world interactions. Evidence from the online survey suggests that participants can use the system based on their answers. However, challenges affecting the system's functionality may occur due to the eye tracker's insufficient accuracy or natural gaze behavior. Participants interacted with the robot to record gaze data, which was then used to evaluate hardware constraints. This study addresses the following questions:

- Would the prototype eye-tracking controller work in the controlled environment of the lab?
- What technical difficulties arise from the design?
- Does perception vary between the hands-on study and the online survey?

Procedure and participants

To minimize the influence of learning effects or behavioral adaptation, 24 participants were involved, creating 24 emerging permutations in the task sequence. Each participant provided informed consent after the procedure was explained. They were asked to sit in a chair in front of the robot workspace. The eye-tracker was set up and calibrated as described in [192]. Participants started the task by fixating on one of two objects in the drinking, eating, light-switching, and picking-and-placing tasks with

gaze (see Figure 5.1). Participants chose the objects according to their preference. The experimenter performed the role of the automation, as described in Section 5.3.2.

Production and recognition variables were discussed prior to the test. Within the first second of selection, the wizard selected the corresponding task via the web interface. The robot's reaction took between one and five seconds from the time the first reaction was recognized, including the wizard's reaction time. As shown in Table 1, one wizard error occurred, causing the robot to stop shortly after the experiment started due to an incorrect mode selection. This experiment's results were discarded, and it was repeated with another participant to ensure naivety to the procedure and function of the Wizard of Oz setup.

Wizard training: The experimenter was informed of the user's task sequence and observed the participant's eye movements in real time on the monitor. As soon as the participant fixated on an object, the corresponding task started via the robot's web interface.

Data analysis

Data were recorded using wearable eye-tracking glasses (Tobii, Tobii Pro Glasses 3, accuracy = 0.6°). Gaze visualization, data analysis, and post-processing were performed using Tobii Pro Lab [193]. In the hands-on study, the computer controlling the robot and visualizing the participant's gaze was placed 2 m behind the participant so that the user was unaware of the experimenter's role. Evaluating the areas of interest (AoI) and dwell time is necessary to estimate the need for error minimization methods and determine constraints for object recognition algorithms. Along with this technical evaluation of the system, we analyzed the perception of the non-disabled participants. In the Tobii Pro Lab events, Times of Interest (ToI), gaze movements during object selection, and Areas of Interest (AoI) were defined and labeled on the data. Tobii Pro Lab calculated metrics such as dwell times, which describe the total fixation time of an AoI relative to the ToI and the number of missed hits, and exported them for further analysis. The technical evaluation was performed based on these parameters.

Interpreting gaze data provides insights into the factors mentioned in Section 5.2.1. ToI, AoI, and dwell time can be used to describe the scene based on visual attention. In order to facilitate the understanding of such terms, definitions according to Holmqvist and Andersson [31] and Tobii [193] are presented below.

- Event: In Tobii Pro Lab, describes the timestamp when an event occurs. In this case, it marks the start and end times of a task.
- Time of Interest (ToI): The interval between two related events.
- Area of Interest (AoI): A polygon that corresponds to the outline of an object. The resulting area describes a visual stimulus.
- AoI hit: The entry of the tracked gaze into an AoI.
- Dwell Time: The time that the gaze remains within an AoI after the initial entry. In this study, the gaze was allowed to exit and re-enter the AoI. Dwell time is the sum of time while the gaze stays within an AoI.
- Type of eye movement: In this study, fixations and saccades were of primary interest. Fixations refer to the time that the eyes remain stable and focused on a specific point or area of interest. They were not distinguished from smooth pursuit, which occurs when the head moves while the gaze remains fixed on a point in the scene. Saccades are rapid changes in gaze direction. They connect fixations. Other eye movements, such as vergence movements and microsaccades, were not investigated. Further information on gaze movement can be found in Holmqvist and Nyström [31].

5.4 Results

5.4.1 Online survey

Of the 71 participants, 36 surveys with no responses (N/A) had to be excluded. 19 participants completed the survey, and 15 partially completed surveys were included when they answered at least Q1 to Q12 (referring to the question indicators stated in Appendix 5.A), due to their given insights into daily life. Therefore, the analysis used 34 complete and partially completed surveys. We assume that the large number of empty surveys resulted from people skipping through the survey out of curiosity. Additionally, some institutions reported that they conducted an internal ethical evaluation before distributing the survey to patients and staff. These surveys were automatically saved, even if they were not finished.

The group distribution of the completed surveys was 10 PSMD, 3 family members or friends (fam./fr.), and 6 medical professionals (med. prof.). In comparison, the in the incomplete surveys 4 PSMD, 6 fam./fr., and 5 medical professionals participated. The demographics can be found in Table 5.1, which summarizes the results of Q1 to Q3. The general and system-related questions varied by group to identify different viewpoints on the system as shown in Figure 5.3. Table 5.2 shows the distribution of questions related to the corresponding topic. The results of Q4 to Q15 will be presented in the following topics.

Support situation

For this topic, questions were asked to evaluate the participants' personal situations and whether there were any obstacles to filling out the survey. Only two of the PSMD group received assistance filling out the survey (Q4(PSMD)). The reason was not stated. Due to limitations in hand and upper body movement, and a lack of access to suitable assistive devices, it can be assumed that the assistance was necessary (Q8, Q9). Relatives and friends only stated once that the person they cared for filled out the survey. Therefore, the risk of bias due to conferring between participants is low.

Level of disability and assistive devices

Q6 revealed that two participants experienced variability in their disability-related limitations from day to day, underscoring the importance of considering individuals with episodic disabilities. These individuals may particularly benefit from multimodal control approaches that can adapt to varying abilities, as discussed by Bhattacharjee et al. and Nanavati et al. [15, 6]. Regarding Q8, participants differed greatly in their motor abilities and, in most cases, their level of disability remained unchanged (Q7: 12 out of 15). In the other three cases, the level of disability improved, which could indicate stroke rehabilitation. This emphasizes that, for this group, the presented control method may only be applicable for a short time before switching to other control inputs, such as a joystick. This emphasizes the importance of systems that are easy to learn and apply, assisting in short recovery times. Relatives and friends stated that three participants had access to assistive devices, in the form of a patient lift and a repositioning aid.

Robotic arm use

Of all the participants asked, only one stated that she had already used a robotic arm controlled by a chin switch. In the context of Q7(PSMD) and Q8(PSMD), it may be that the level of disability of the other participants does not qualify them to receive a robotic arm from healthcare providers, or they may receive assistance from caregivers or family, which is more common in Germany, Switzerland, and Austria. Approval for a medical device by healthcare providers in Germany can take several months

or, in rare cases, years. This indicates that the use of robots is uncommon and can greatly bias the participants' perception of the system due to its novelty.

WoZ satisfaction

After viewing the video, the participants were asked if they or the person they cared for would be able to use the system. More than half of the participants confirmed this (17 out of 34; Q13 - PSMD, Q14 - Rel./Fr. and Med. Prof.). Only three participants stated the opposite. No reasons were given. Medical professionals and family members were also asked if they were satisfied with the setup process (Q11 - Rel./Fr. and Med. Prof.). Five stated they were satisfied, while three were not. One reason given was the amount of time it took to set up. It might also be influenced by the desire for more information about the system. As Figure 4 shows, the high number of N/A answers in the partially completed surveys occurred after the video. This may be because participants were unable to return to the website after watching the video. Due to missing feedback, this is only a presumption.

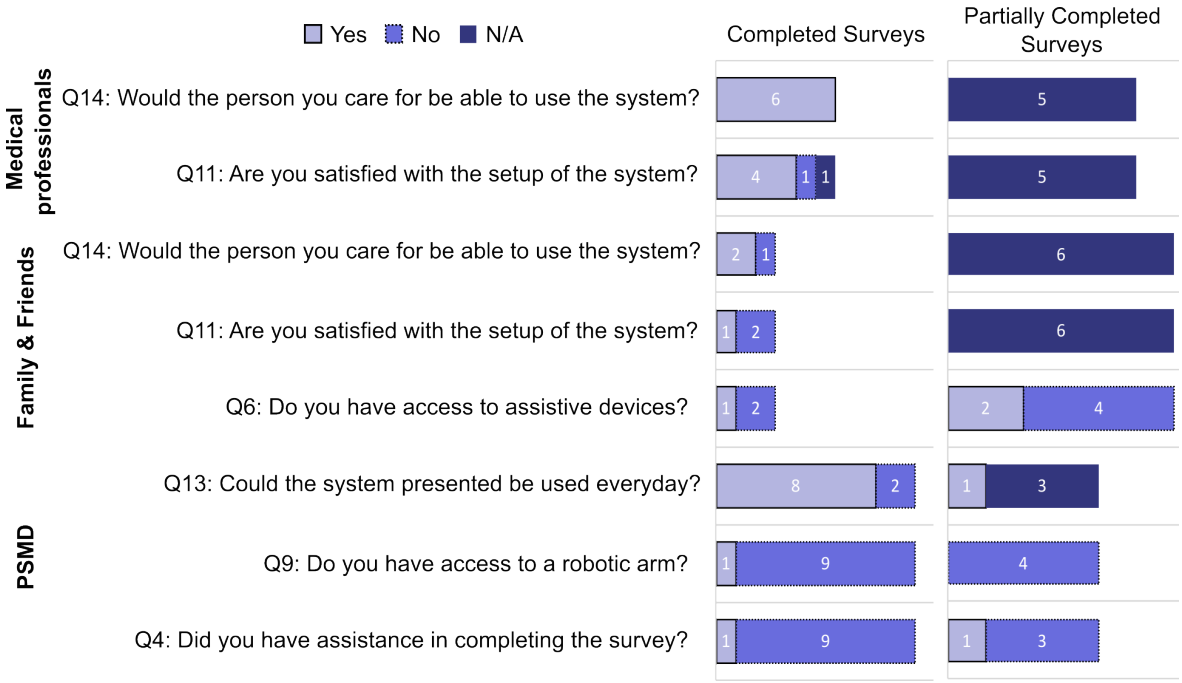


Figure 5.4: Distribution of responses from online survey participants. Q4, Q9 and Q13 were answered by PSMD, Q6, Q11 and Q14 by family and friends and medical professionals.

Task preference

Participants were asked what tasks they would prioritize to accomplish with the system. The group of PSMD gave a total of 14 responses: Eating (3 mentions), drinking (2 mentions), reaching for cutlery, pushing buttons (elevator or electric door), applying makeup or brushing their hair, cleaning, scratching, opening a bottle, and lifting objects of various weights from the floor (2 mentions). Family members and medical professionals were asked this question twice. First, they were asked to identify situations in which a robotic system could assist with daily tasks, while they were unaware of the system demonstration. Second, they were asked about the presented system. Thirteen people responded to the first question. They rated the following tasks as useful: Making the bed, brushing teeth, assisting with eating, thrombosis prevention, putting on socks, documentation, going to the bathroom, doing laundry, cleaning (2 mentions), assisting with walking, assisting with doctor's appointments and correspondence,

silent alarm for service calls, and making phone calls. Fourteen responses were given to the second question. Eating (5 mentions), drinking, turning on lights, operating the aforementioned lift system, brushing teeth, preparing a drink, opening doors, lifting objects (2 mentions), making phone calls, and moving objects. Both questions demonstrate a high potential for robotic systems and digitization. The first focuses on assistance with household tasks and communication, while the second focuses on user-centered tasks. Figure 5.5 shows the distribution of task preferences mentioned by the participants.

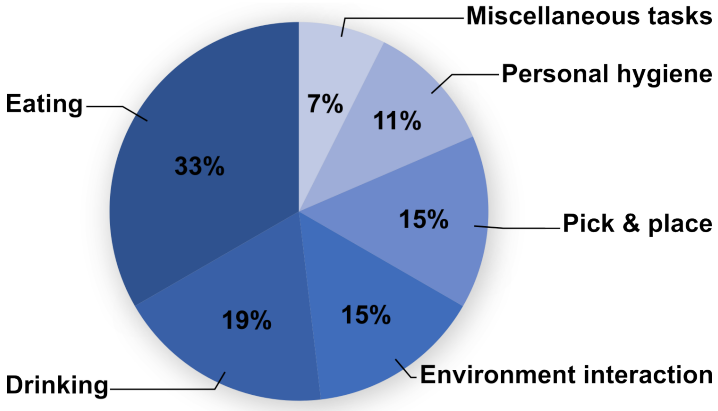


Figure 5.5: Classification of all mentioned tasks in relation to the presented system. 28 tasks were classified into six categories containing the tasks mentioned by the participants.

5.4.2 Hands-on study

Gaze behavior analysis

Analyses of dwell times in relation to object selection events were used to determine the technical requirements for the eye-tracking controller. Assumptions were investigated, such as the correlation between object size and dwell time duration.

Table 5.3 shows the distribution of participants' object selection choices and dwell time information. A visual attention map is shown in Figure 5.6. The chosen interval was 1,043 ms before the robot started moving, based on the time steps made by the evaluation software. Both data in Table 5.3 and in Figure 5.6 represent the intervals. During this time period, events such as searching for the object or deciding which object to select were excluded. Participants were instructed to focus solely on the selected object, to investigate whether natural gaze behavior exhibits gaze cues that hinder implementation or result in insufficient accuracy. The average duration of the test was 2 min, 41 sec (SD = 9.8 sec).

The average dwell time was calculated to determine the maximum limit for the controller's dwell time, including the time from the start of fixation until the robot moved. In this setting, the longest possible dwell time for the participant's gaze was found to be 5.612 seconds. During this time, the gaze did not leave the area of interest (AoI). This indicates that dwell time triggers can be set to long durations. However, the following limitations show that such long dwell times are rare.

In this study, we assumed that dwell time would be longer for larger objects due to larger AoIs. Accuracy errors would have less impact on AoI interactions than in the small AoIs. The correlation between object sizes and corresponding dwell times was measured and is given in Figure 5.7. The dwell time of the bottle has a higher standard deviation than the cup, implying more occurrences of longer dwell times. As seen in Table 5.3, this is also visible in the dwell times above 500 ms. Reasons why the correlation may not apply are the fixation of the bottle cap, which is closer to the edge of the AoI than the cup handle. The gaze is more likely to leave the AoI, resulting in lower dwell times, which

indicates the influence of *user intention*. In comparison, the fork and the block have similar average dwell times and standard deviations, despite their different sizes. We extended our initial hypothesis to include the influence of *geometry of the object* on the dwell time. We tested this by evaluating missed hits and gaze behavior with a fork handle that was less than 1 cm wide. When accuracy errors occur, it is more likely that gaze is tracked outside the fork's AoI.

Table 5.3: Information about distribution of selection by participants, dwell times and missed AoI hits.

Object	Number of selections	Occurrences of dwell time over 500 ms	Missed AoI hits
Bottle	11	82%	0
Platform	7	87%	1
Cup	13	92%	0
Light switch	24	75%	4
Fork	9	58%	5
Block	14	56%	5
Food item	11	38%	7

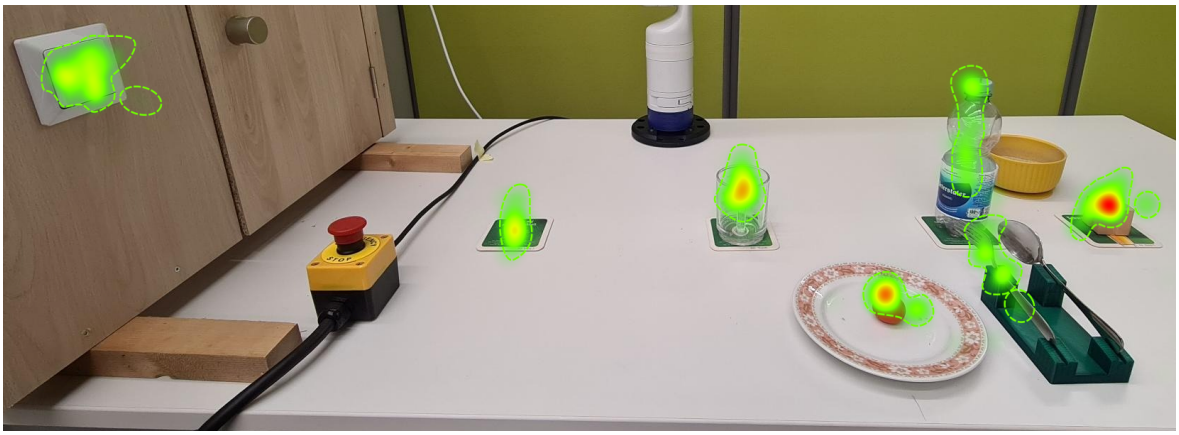


Figure 5.6: Heat map visualizing the attention of participants during task selection (absolute duration), mapped to the snapshot of the experimental setup. Red areas indicate frequent fixations, green less fixations. Areas without coloring indicate no fixations within the observed intervals. Dashed outlines were added manually to better distinguish the edges of the hotspots.

Examining the low dwell times of the food item placed closest to the user revealed the influence of *gaze direction* on the dwell time. Looking at the gaze distribution in relation to the scene camera's coordinate system reveals that participants tended to move their heads to center their gaze when looking at objects to the right or left of them, as shown in Figure 5.8. In contrast, participants tended not to move their heads when looking down. The detected gaze location was less accurate due to the greater scatter closer to the eye-tracker's periphery. Physiological factors, such as shadowed pupils by the eyelashes, cannot be excluded to impact dwell time. This resulted in seven missed AoI hits for the food item.

Significant correlations were investigated between dwell time, object size, and gaze direction towards the object. The Shapiro-Wilk test ($H = 0.771$, $df = 97$, $p = <0.001$) and the Levene test ($H = 2.658$, $df1 = 6$, $df2 = 90$, $p = 0.02$) showed no normal distribution or homogeneity. A Kruskal-Wallis test was applied and showed significant differences ($H(6) = 17.442$, $p < 0.008$). Post-hoc analysis using Dunn's test with Bonferroni correction revealed significant differences ($p < 0.05$) between the following pairs: food item and bottle, food item and platform, food item and light switch,

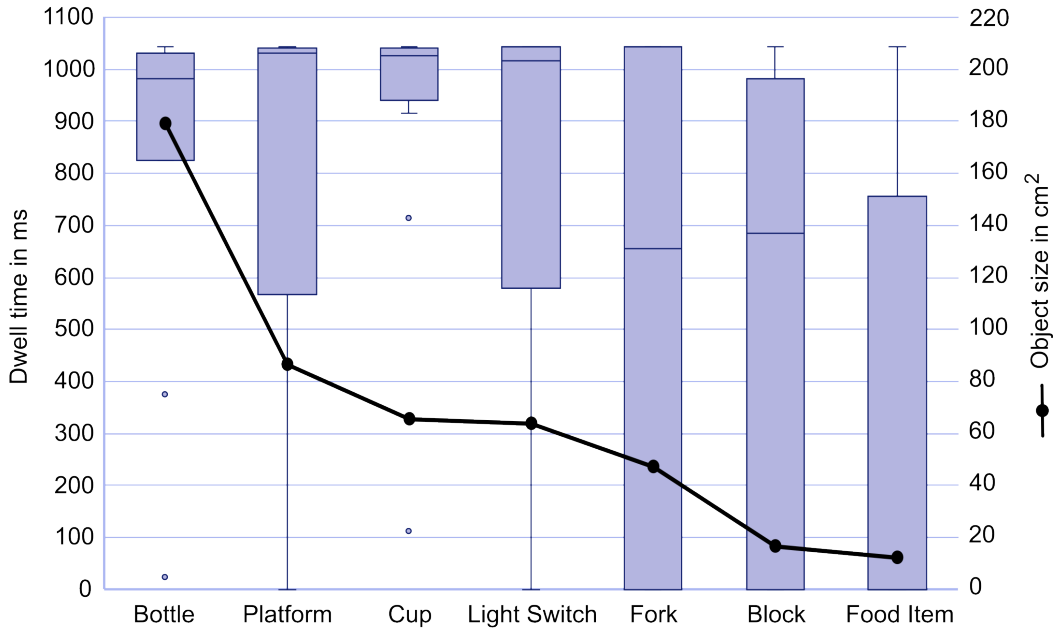


Figure 5.7: Correlation between object size, missed hits and dwell time for the selectable objects. Dwell time is given as the total duration of gaze within the objects AoI 1043 ms before the robot started moving. The line in the box plots marks the median, confidence intervals are given at 95%.

food item and cup, and block and light switch. These results indicate that the measured data for the food item are significantly different from the other objects. Spearman's ρ was used to analyze the correlation between object size, object position, and dwell time. In Figure 5.8, gaze location was evaluated by applying a coordinate map visualized by the equally distributed sectors. The analysis showed that dwell time and object size ($\rho = 0.373$, $p < 0.001$) were positively correlated, as well as dwell time and offset in y-direction ($\rho = 0.263$, $p = 0.009$). The correlation between the offset values in y-direction and x-direction is weaker and does not correlate with dwell time ($\rho = 0.160$, $p = 0.116$). This indicates that the assumption that the dwell time is influenced by *gaze location* towards the object and the *object size* is correct.

Perception of hands-on application

In general, the hands-on study revealed satisfaction with the system. In the unstructured interview, the participants' comments were evaluated based on whether they were positive or negative regarding the robot's performance. The results were as follows: nine positive comments, three negative comments, nine comments that were both positive and negative, and three comments that were neither positive nor negative. All statements can be found in Table 5.6 in Appendix 5.B. Differences were found between the responses of people who regularly work with robotic arms (the prior-experience group) and those who have no experience with robotic arms and robots in general (the non-prior-experience group). The non-prior-experience group gave mostly positive comments and focused on describing the tasks, while the prior-experience group focused on robot kinematics, robot behavior, and user safety.

The non-expert group reported feeling interested, entertained, satisfied, and impressed. At the same time, they reported feeling confused and uncertain. The task design itself caused confusion because participants were unsure how to react to the glass presented by the robot. Considering a real-life application of pouring a drink, one participant expressed fear of overfilling the glass. Regarding the robotic system, participants were satisfied with good reaction time and intuitive control. Negative

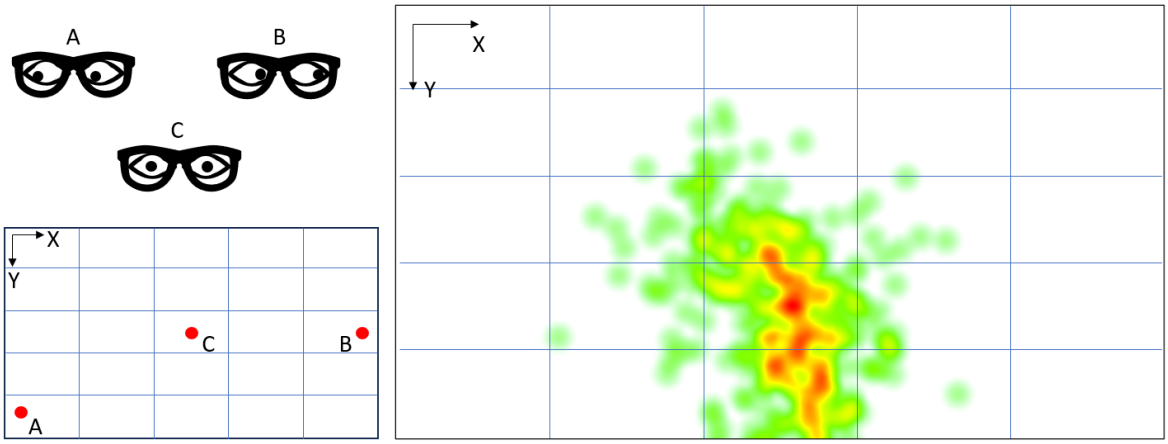


Figure 5.8: Left: Exemplary representation of the viewing directions to explain the resulting heat map. Visualization of the viewing point in relation to the visual field of the scene camera. Right: Heat map of gaze position during task selection. The field of view is divided into 5 by 5 fields indicating a 5-section coordinate map used to determine the correlation between dwell time, gaze location, and missed AoI hits. Gaze direction was mostly centered and downwards while task selection.

comments concerned the velocity, which created ambivalent reactions. Some participants stated the velocity was just right, and others saying it was too slow. Importantly, users reported feeling safer because the robot's movements were limited to the edge of the table. The fork movement toward the participant had an unsettling effect, as they reported that the fork was not close enough to eat, or at the right height for their mouth.

Comments from the prior-experience group were mostly negative about the robot's trajectory and the measurement setup. For instance, the gripper could potentially crush someone's hand if the user reached into the setup due to the lack of pressure sensors. The placement of the emergency stop was a concern for the well-being of future users, due to its distance from the participant. They also mentioned the robot's shaky movements, non-optimized trajectories, and one instance where the robot accidentally touched the bottle while moving to the next object. The perceived motor vibration was uncomfortable and the sound of the dropped fork was too loud. Reaction time and object recognition were mentioned as positive, implying that the Wizard of Oz design was convincing.

Additionally, two participants from both groups stated that they were uncertain whether the robot would stop the task if they looked away or at other objects while the robot was performing it, a concern that must be addressed in the control design.

5.5 Discussion

In Section 5.5.1, two of the three usability questions and in Section 5.5.2, two of the three functionality questions are answered. In Section 5.5.3, the last question of both studies will be discussed, which combines the perception of the online study and hands-on study. Feature improvements to the initial design are presented.

5.5.1 Usability

Are potential users able to use the system?

Yes, the survey showed in Q13(PSMD) and Q14(Fam./Fr. and Med. prof.) (Figure 5.4) that most of the participants are able to use the system based on the video presentation (Answers: 17 yes, 3 no, 14

N/A). The high amount of N/A results were the result of including partially filled out surveys.

The limiting factors that prevent the system from being used every day were not stated. In the open-ended questions, participants stated that they would like more information. They also said that they would like more information about the system to better evaluate its maintenance and function. These requests did not imply that the participants would change their minds after receiving the additional information. One participant stated, “I wonder if the system can be used by people with intellectual disabilities due to their reduced cognition”. This highlights the importance of considering various impacts of disabilities during the design process to make such systems accessible to a wide range of users. It again highlights the necessity of involving the community in the design process.

Deeper issues, such as how the control works and whether wrong selections can occur due to natural gaze behavior, were not questioned by the participants, contrasting the results from the hands-on study. Section 5.5.3 shows participants’ concerns about interacting directly with the system, highlighting the benefit of dual investigations in robot perception.

Which tasks should be prioritized to maximize the robot’s usefulness in everyday life?

The results, shown in Figure 5.5, the most requested tasks are assisting while eating or drinking, followed by environmental interaction, which includes turning lights on and off, pressing buttons, opening doors, and picking and placing objects. Personal hygiene consisted of a variety of different tasks, including doing makeup and hair, brushing teeth, and cleaning. Miscellaneous tasks included making phone calls and scratching themselves.

Brushing one’s hair, applying makeup, or scratching are daily tasks that require close human-robot interaction while ensuring the safety of the user. The shared control approach presented would require face and body recognition if the user is unable to move their head to a certain position. Ethical questions arise when personal data is recorded, including physical and physiological appearance. Current discussions in this area suggest that the user must be informed about the system functions and the personal data collected. This would allow the user to decide whether the privacy limitations or the benefits of the system outweigh [155]. The aim must be to present this information in simple language so that people with physical and cognitive disabilities can understand it and make decisions.

The prioritized tasks and the tasks realized from the literature included in Section 5.2.3 were summarized. A comparison of the results with the findings in the literature shows partial agreement. Table 5.4 compares findings from the literature with our results. In the "Prioritization" column, we summarized the works of Beer et al., Nanavati et al., Petrich et al., Smarr et al., Stanger et al., Chang et al., and Jardón et al. [172, 4, 6, 173, 176, 189, 188]. These works present or review task prioritization in activities of daily living (ADLs), involving either elderly people or people with disabilities. It is important to acknowledge that these works include varying participant groups, as well as varying robot types. To the best of the authors’ knowledge, no review or work solely focuses on PSMD and robotic arms. The "Integration" column summarizes reviews and individual studies concerning the execution of tasks with assistive robots. It summarizes the outcomes of the following works, which were also presented in Section 5.2.3 [6, 5, 181, 65, 22, 26, 163, 15]. Some tasks were excluded from the literature when they could not be performed with a robotic arm (e.g., getting around or into bed, going out, toileting, managing money, leisure and recreation, and shopping), as well as outdated tasks (e.g., removing paper from printers and fax machines).

Tasks such as preparing meals, dressing, bathing, grooming, and taking medicine were not mentioned by the participants. This may be because these tasks are comparatively less important to end users, as found in the literature. In contrast, preparing meals is highly prioritized and researched in other studies. Depending on their hobbies or the type of personal assistance they receive, this task may not have been favored by the small group of participants. The insight into the eating task is interesting. While most participants stated that they strongly preferred to accomplish this task independently,

Nanavati et al. found that, compared to other ADLs and IADLs, this task was the least prioritized [6]. However, other sources, such as Petrich et al. and Stanger et al., found it to be highly prioritized [176, 4]. This discrepancy may be due to the fact that Nanavati et al.'s work reviewed different populations, including people with visual impairments and the elderly, who may be able to eat without assistance.

Comparing works that address these tasks reveals gaps in interacting with switches and buttons, personal hygiene, scratching, and using electronics. Concerns about user safety and privacy regulations might influence the low number of works related to personal hygiene and scratching. In contrast, using electronics or phones is becoming more convenient with smart home applications and hands-free technology. In conclusion, the chosen task for the presentation aligns well with the literature. More focus should be placed on assisting people with personal hygiene tasks. The variety of tasks performed demonstrates that the participants were not influenced by the tasks shown in the video.

Table 5.4: Overview of tasks mentioned by participants and representation within literature. Tasks mentioned by users are presented to their according categories in Figure 5.5. Fit describes if the task can be accomplished without additional assistive devices to a robotic arm (triangle shows partial accomplishment). The comparison is done between our results, task prioritization found in other works and integrated tasks in existing works.

Category	Tasks	Fit	Our results	Prioritization	Integration
Eating	Eating	✓	↗	→	↑
Drinking	Drinking	✓	↗	↗	↑
Environmental Interaction	Preparing meals	✓		↑	↑
	Switches, Buttons	✓	↗	↑	→
	Doors, drawers	✓	↗	↑	↑
	Pick and place	✓	↗	↑	↑
	Carrying objects	✓	↗	↑	↗
Personal hygiene	Makeup, shaving	✓	↗	↑	→
	Combing hair	✓	↗	↗	↘
	Brushing teeth	✓	↗	↗	↘
	Dressing	✓		→	→
	Bathing, Grooming	△		→	↗
Miscellaneous tasks	Using phone, electronics	△	↗	→	→
	Taking medicine	✓		→	↘
	House choirs	✓	↗	↑	↑
	Scratching	△	↗	↗	↘

5.5.2 Functionality

Would the prototype eye-tracking controller work in the controlled environment of the lab?

Initial testing showed that 70% of object selections would result in selection success when the dwell time trigger is set to 500 ms. Although dwell times can be shortened to enhance selection success, as shown in other literature [31], this could lead to an unintentional object selection, as described in Section 5.2.1. An appropriate threshold must be explored with the prototype system.

Accuracy issues are one of the main reasons why gaze was tracked outside of an AoI. Therefore, the influences on AoI misses and dwell time were investigated in Section 5.4.2. Four dependencies were found:

- Object size: As shown in Figure 5.7, object size correlates significantly with dwell time. Small objects, like keys, are part of everyday activities. Excluding such objects from interaction would lead to major drawbacks in participation in daily life. The available workspace is limited by the length of the robotic arm. This also limits the relative size of the object in the camera scene. With this limit, the resulting accuracy error can be calculated and checked with the desired interactive object size to estimate if this would cause problems. In this case, the accuracy of the eye-tracker is $\Delta_{deg} = 0.6°$. With an effective workspace of $d = 89.18$ cm for the used robot arm [194], the maximum accuracy error is 9.34 mm ($\Delta = d \cdot tan(\Delta_{deg})$). Since all selected objects were larger than 9.34 mm, it is assumed that the prototype would work.

- Geometry: The AoI of the fork could lead to a failed selection due to its width. Currently, there are two approaches in planning for estimating object outline. Object recognition can be used to create bounding boxes, solving the issue due to the new geometry. Overlapping of bounding boxes may occur more often, leading to unintentional object selection and context misinterpretation. Object segmentation can lead to boundaries which fit the object's outline. In this case, the geometry of the object would be traced with high accuracy, solving the bounding box issue, leading back to the initial width issue.

- Gaze location towards the object: In the eating task example, participants tend to move their heads to the left and right instead of up and down (see Figure 5.8). The resulting higher accuracy error was due to the glasses periphery. Considering physical disabilities of potential users, an immobility of the head cannot be excluded.

- User intention: Participants were instructed to focus their gaze on the selected object, but not on which region in detail. In the case of the bottle, the task was to fill the glass. The bottle needs to be opened first, which explains why the bottle cap was fixated, indicating the hand-eye coordination and intention behind the task, as discussed in Section 5.2.1.

What technical difficulties arise from the design?

Ambiguous object selection can occur due to object overlap and was found in the case of the fork (see Figure 5.6). In addition to the accuracy challenges, robust object selection may be the most difficult challenge in implementing the prototype. Other researchers have addressed this issue using 3D gaze tracking to minimize localization errors in approaches involving eye-tracking glasses [10, 20]. As discussed in Section 5.2.1, user intent may lead to more robust selection methods by involving multiple objects to estimate task intent. These objects may vary in geometry and size, which reduces the impact of incorrect fixations. Regarding current approaches using object detection, overlap in bounding boxes of closely arranged objects may lead to larger issues than those found in the area of interest (AoI) evaluation. The AoIs were fitted to the outlines of the objects, similar to object segmentation. However, bounding boxes have larger areas, which leads to ambiguity when they overlap.

Section 5.2.2 presents technical challenges, level of autonomy, and user preferences as factors influencing user acceptance. Within these categories, we found indications from the survey and hands-on study of technical difficulties.

After implementing the suggested control method, the tasks will be selectable as planned. However, as Hagengruber et al. [5] point out, accomplishing subsequent tasks is necessary to realize benefits for everyday usage. With this in mind, objects could indicate multiple tasks. Solutions integrating machine learning methods to identify gaze patterns, as presented by Yang et al., Jain et al., and Belardinelli, can reduce misinterpretation [20, 89, 11].

In the WoZ design, the robot was presented with a high level of autonomy. As discussed in Section 5.2.2, it is questionable whether participants will accept the system as intended. In the hands-on study,

participants vocalized their safety concerns, especially in the group with no prior robot experience. Arbitration methods can improve perception. This is especially important, since robot perception changes with regular use, and the level of autonomy can be reduced until the user is confident enough to operate the robot with a higher level of autonomy. Additionally, settings should be integrated that allow users to adjust the robot's velocity to their preferred level.

5.5.3 Differences between Online Study and Hands-On User Study

Were the participants satisfied with the presented system?

In both studies, the participants were mostly satisfied with the system. We found in both studies suggestions for improvement. The design can be refined to meet the participants' expectations of the system. The framework must have the following features to realize the system:

- Gaze detection on objects: Due to the outcomes in Section 5.5.2, several approaches are followed to detect objects in the scene camera of the eye-tracking glasses. The prototype will estimate unintentional selection and context misinterpretation.
- Task availability: The survey showed that the desired tasks were in line with the participants' expectations. Eating was the most frequently mentioned task. However, handling a fork is a safety risk when the robot manipulates the fork autonomously, and has to be handled appropriately.
- Task selection: The expected prototype will implement objects associated with multiple tasks. Testing has shown that if the dwell time is set correctly and the algorithm is designed to wait for multiple objects to be selected, the task should be robustly selectable. Dwell times of 500 ms and less were found to be appropriate. As mentioned in Section 5.2.1, feedback strategies need to be evaluated and integrated into the control to ensure the user. Strategies have to be evaluated if user intent methods assist in solving issues selecting between multiple tasks.
- Trajectory, grasping, and collision avoidance: The robot has to perform the task by itself. In the current version, the robot's trajectory was programmed, and the objects were placed at the same locations for each test. In a real-world scenario, this would not meet the user's expectations. Solutions were found, as presented in Section 5.2.1), and depend on the approach chosen for object gaze detection.
- Adaptation: Due to the ambiguous results of the robot's velocity perception, settings will be included in the prototype so that users can adapt the system to their liking, such as robot velocity.

In future work, the experimenter's control has to be automatized so that the robot can autonomously perform tasks selected by the user.

Does perception vary between the hands-on study and the online survey?

The unstructured, in-person interviews with participants in the hands-on study and the open-ended questions in the online survey provided more diverse feedback on this shared control design. The results are consistent with, and extend, current literature in this area. It adds information for the design of gaze-controlled assistive robotic arms. Perceptions of the system do not vary, but highlight different aspects of the design.

Involving the stakeholder group yielded insights about daily tasks and their daily life with their disability. The prior-experience group revealed potential drawbacks in the technical design process, such as concerns about safety, kinematic selection, and trajectory planning. Although the non-prior-experience group did not have in-depth feedback on certain aspects of life with a disability or technology,

they helped build a larger pool of participants to measure gaze behavior. In addition, their perception of the robot was more curious and, at the same time, hesitant than that of the prior-experience group due to the novelty of robot interaction. Since robotic arms are rarely available to PSMD, it is reasonable to assume that, in a real-world test, PSMD would respond in an initial robot interaction similarly to the non-prior experience group.

5.6 Limitations

Online study: The qualitative analysis of the online survey provided limited insight into the participants' everyday lives, meaning the data cannot be generalized to a larger population. To counter this issue, we enriched and compared our information with suitable literature. For future studies, however, we have decided to conduct interviews with participants to minimize potential hurdles due to a lack of assistive devices or display problems on end devices. Q4 was tailored to evaluate whether hurdles exist. No indications were found. Additionally, no emails stating technical hurdles were received from medical professionals, family, or PSMD. The online survey might have benefited from additional questions. The safety concerns were not reflected in the survey responses. Including questions such as "Would you feel safe interacting with the robot?" would have provided a more comprehensive understanding of this topic. However, other studies show that potential users are more likely to accept the risks of using the system than reject it, becoming dependent on medical professionals or family members [157]. This can also be seen here, as the group with robot experience is aware of the risks of human-robot interaction, but is not fearful of interacting with the robot. The survey did not ask about the occupation of the medical professionals. Therefore, opinions may differ depending on the occupational field (e.g., doctors, caregivers, and nurses). The online survey was designed for quantitative analysis. Due to the limited number of participants, a qualitative analysis was performed. Therefore, not all answers were discussed in detail, since they are not representative (e.g., Q5 rel./fr.: time spent on care). These results can be accessed in the supplementary materials. As discussed in Section 5.3.3, participants did not interact directly with the robot, which might affect perception.

Specific Wizard of Oz design: The Wizard of Oz system does not vary in intent interpretation due to its hard-coded trajectories. Variable behavior was excluded from this study, since comparing outcomes between participants would not be applicable. However, varying behavior in shared control changes the perception of the system, as shown in [90]. We presented a design driven by eye-tracking, in which objects are selected without the need for a GUI. This approach is uncommon, since prototype systems face challenges such as interpreting the desired task, estimating errors due to head movements, and lacking feedback modalities [12]. If differences in control design exist in relation to the WoZ design, the results may not be transferable to other controls.

Generalized approach: Since this approach was not tailored to specific disabilities, it may lack the details necessary to adapt to the needs of individual users. Future work can use this generalization to more easily identify these requirements.

Transfer to other eye trackers: Only the Tobii Pro Glasses 3 eye tracker was used, which is more accurate than other eye-tracking glasses on the market. However, the effect on the bounding boxes in object detection can vary greatly depending on the model used. Due to the scope of this paper, the test was not reproduced with other eye trackers.

5.7 Conclusion

The goal of this work was to obtain feedback on an eye-tracking-driven shared control for assistive robotic arms in order to identify community expectations, desired tasks, and technical challenges. While previous studies offer valuable findings, they vary between robot types and involve alternative

stakeholder groups. As a result, the available insights are generally only limited transferable and not specific to the control for an assistive robotic arm examined here. As suggested by other researchers, a Wizard of Oz design was created prior to prototyping. The factors necessary for successfully implementing shared control were identified and presented.

Based on this WoZ design, two studies were conducted. An online survey designed for PSMD, their families and friends, and medical professionals. This survey provided feedback on the community's expectations for the system and the tasks the robot should be able to perform. A hands-on study with participants with and without robotics experience provided insight into the technical requirements and feedback differences between the video presentation available for the online survey and the real-world application. Gaze data was recorded to measure dwell times and influence dependencies related to the eye-tracking controls.

The duality of both studies provided insight into the usability and functionality of the system. Task prioritization was contextualized with current literature and approaches that researchers presented to realize shared controls. Dependencies were uncovered for the design of eye-tracking-driven robot controls, such as the influence of object size, geometry, gaze location, and user intent on accuracy. The impact on object detection or online task intent recognition based approaches were assessed and discussed. Consequently, improvements and open challenges related to the initial idea were presented.

5.A Overview of Questions asked in the Online Survey

In the following an Overview of the questions asked to the participants in the online Survey are presented. They represent three versions of the survey which would be handed to the individual participant group PSMD, family and friends and medical professionals. All related answers can be found in the supplementary materials.

Table 5.5: This table presents all questions asked in the online survey. Questions spanning over all three columns indicate questions that were asked to each participant.

Indicator	Questions asked to group		
	PSMD	Relatives and friends	Medical professionals
Q1	What age group are you in?		
Q2	What gender are you?		
Q3	To which of the groups below do you feel you belong?		
Q4	Do you have assistance from another person in completing this questionnaire?	Did the person you care for participate in this survey?	Please estimate the percentage of your patients or clients who have movement difficulties in their arms?
Q5	If so, what is the reason you need support?	How much time do you spend on care each day (in hours)?	Let's assume that an assistance system could take over all daily tasks. Which tasks would you most like to hand over in order to facilitate care? Name 1-3 favorites.
Q6	How often does your impairment affect your daily life?	Do you already use assistive devices to help you provide care?	-

Q7	In terms of your physical disability, has your condition improved, worsened, or stayed the same in the last 6 months?	If yes, which assistive devices do you use?	-
Q8	Select the body movements that you can control without restrictions	-	-
Q9	Do you already use a robotic gripper arm, for example on an electric wheelchair?	Does the person(s) you care for already use a robotic gripping arm, e.g. on an electric wheelchair?	
Q10	If so, how do you control it?	If so, how does the person control this robotic arm?	
Q11	Are you satisfied with your current system?	Are you satisfied with the system in terms of startup and function?	
Q12	If no, indicate why you are not satisfied with the system.	If no, why are you not satisfied with the system?	
Q13	In relation to your disability, would it be possible for you to use the system presented in the video every day?	In which tasks would you or the person you care for most like to be supported by this system in daily care? Please name your 1-3 favorites.	
Q14	If no, what changes would you suggest to make the system easier for you to use?	Would it be possible for the person you care for to use the system presented in the video every day?	
Q15	In which tasks would you most like to be supported by this system? Please name your 1-3 favorites.	If no, what changes would you suggest to make the system easier to use for you and the person you care for?	
Q16	Do you have any other comments on the questions asked or the system?		
Q17	Would you like to have had more detailed information on how to use the system to answer the questions asked here?		
Q18	If yes, what information would you have liked to have?		

5.B Participant statements of hands-on study

In the following the statements of participants of the hands-on study are listed. All statements were translated by the authors to english. Participants without any statement to the system were not listed below.

Table 5.6: Statements of participants. Prior-experience group: Exp., Non-Prior-experience group: Non-Exp.

P	Exp./ Non-Exp.	Statement
1	Non-Exp.	By noting that the robot does not go over the table, P1 was reassured. “apprehension, causious” system if not had stated, Glass’s task: fear of overspilling and dropping the cup in the real task.
3	Non-Exp.	Food task: Lifting a fork so high that you only had to grab was cool. Light switch was hit centered, that was cool. The handle of the cup was held towards the person. That was very practical if you wanted to grab it.
4	Non-Exp.	Fork task: Test person was confused because the fork pointed to a fixed point and did not go in the direction of the mouth. General: Confused as to whether the robot cancels the task when looking at another object.
5	Non-Exp.	Movements too slow for everyday use. However, it was satisfactory for an initial test.
6	Non-Exp.	Food task: Serving food was interesting. Glass task: The glass slipped a little when I put it down. There was no straw in the glass, so it wasn’t clear why the robot handed the glass flat.
7	Non-Exp.	Open gripper is uncomfortable when it points at you. Velocity was good.
10	Non-Exp.	General: By saying that the robot does not reach over the table, the action was not scary.
11	Non-Exp.	General: Responds quickly, which is good. Intuitive operation. Once you know where the robot stops it is less intimidating/scary.
13	Exp.	Fork task: First close the gripper a little before gripping the fork completely, less risk of crushing the hand. General: Emergency stop is poorly positioned.
14	Exp.	Kinematics moves spongily, drinking task: Kinematics has touched the bottle.
15	Exp.	When the fork is dropped, a loud unpleasant noise is heard from the drop.
16	Exp.	P16 asked whether tasks are aborted if you accidentally look at other objects. Waiting time is appropriate until the robot moves. It does what you want (intuitive control). Kinematics vibrate during movement, which is negative.
17	Exp.	Robot is shaky, jerks. Selection process: Fear of triggering the wrong action when looking at the wrong object.
18	Exp.	Movement was a little too fast to the light switch
22	Exp.	With pick and place, it was assumed that either the block or the platform could be gripped. Accordingly, P22 was confused.
23	Exp.	Robot trembles, trajectory not optimal, bottle badly twisted while handled.
24	Exp.	Fork was recognized well, even though you looked briefly at the bottle during the time you had to aim. The pause between aiming and starting the pick and place task was too long.

Chapter 6

Publication IV: Robust and Scalable Task Selection for Humans and Robots with the Use of Pictograms as Fiducial Markers

Summary: The control strategy presented in this work was motivated by the results of the Wizard of Oz (WoZ) studies discussed in Chapter 5. The WoZ design studies revealed that selecting from many different tasks in unstructured environments could result in an insufficiently robust task selection process due to natural gaze behavior, bounding box overlap, or accuracy errors that extend beyond the boundaries of the selectable area of interest (AoI). Therefore, the strategy was adapted to select tasks with pictograms fixed to adjacent objects. Pictograms are advantageous because they are easy for caregivers to apply and straightforward for users to understand and distinguish. It is assumed that selecting tasks with pictograms will reassure users more than selecting tasks by gazing upon the object itself, thereby maintaining a sense of agency. Pictograms can serve as fiducial markers, providing supplementary information about the corresponding task, such as the required object classes or the object's location.

The study presented an image dataset and a model trained to detect pictograms. The discussion covered the system's usability and accessibility in detail, as well as its key contributions to robust, scalable, and user-friendly task selection. These insights incorporated the literature review from Chapter 3, as well as works from research areas such as user intent recognition and the use of AI in robot controls. A key finding was that the model's performance affects the robustness and scalability of the selection process. In the event of performance deterioration due to environmental changes, pictograms may be misinterpreted or not detected at all. These findings motivated the development of a fallback mechanism for eye-tracking-driven controls, ensuring correct functionality and robustness in such cases. The model was implemented in the eye-tracking-driven control in Chapter 7.

Contribution: Anke Fischer-Janzen was the primary author of this publication. She was responsible for its conception, implementation, evaluation, and manuscript preparation. Thomas M. Wendt and Kristof Van Laerhoven contributed to the work through supervision, methodological discussions, and manuscript review.

Robust and Scalable Task Selection for Humans and Robots with the Use of Pictograms as Fiducial Markers

Anke Fischer-Janzen
Offenburg University of Applied Sciences
Anke.Fischer-Janzen@hs-offenburg.de

Thomas M. Wendt
Offenburg University of Applied Sciences
Thomas.Wendt@hs-offenburg.de

Kristof Van Laerhoven
University of Siegen
kvl@eit.uni-siegen.de

Abstract

Controlling robots through eye-tracking enables people with severe physical disabilities to perform everyday tasks independently. Interpreting the intent of users from their gaze alone, considering the presence of multiple objects and diverse tasks, poses a significant challenge that influences the design of this control modality. The objective of error-free object or task selection, which enhances user safety, stands in contrast to the improved usability resulting from increased robot autonomy. This paper presents a task selection method that employs easily comprehensible pictograms as fiducial markers for robot control. The pictograms represent eight daily tasks that can be performed with a robotic arm. We present a pilot dataset that contains 1,008 in-the-wild images. The performance of a YOLOv12n model, which was trained on this dataset, was evaluated through a stress test to identify shortcomings in its real-world performance. The design of the pictograms was evaluated for their comprehension in an online survey, resulting in a mean correct response rate of 83%. Based on these results, we discuss the usability, accessibility, scalability, and robustness in the context of the anticipated application case. Our method provides a robust and scalable way to select from a variety of daily tasks that can be accomplished in domestic and public environments.

Keywords: Assistive robots, Assistive technologies, Gaze Tracking, Object detection, Robust control, Scalability

6.1 Introduction

The level of autonomy exhibited by shared controls varies among different robotic applications, including robotic arms, social robots, and humanoid robots [177]. It empowers people with physical disabilities to achieve independence by assisting them with daily task performance. In these shared controls, task execution is delegated to the robot to reduce the cognitive workload of the user[10]. One of the ongoing challenges is achieving seamless and intuitive interaction between humans and robots [16]. Gaze has proven to be a promising input modality for realizing Human-Robot Interaction (HRI) in shared controls. In this research domain, a prevalent methodology is to use gaze as an explicit input modality by fixating gaze to directly point to objects, implying the intention to execute a designated task. Engaging with a tangible environment or using a display can achieve this objective [12]. In contrast to gaze gestures, which are predefined eye movements that instruct the robot, this method requires less cognitive effort and reduces user fatigue [161]. In real-world applications, a single object may be utilized for multiple tasks, creating ambiguity in the selection of the intended task. For example, users could intend to place, fill, or drink from the same glass, showing that the gaze-pointing method would not be applicable.

Observing gaze behavior in multi-object scenes has yielded encouraging outcomes for detecting user intent in the context of task prediction [13, 65, 22]. Observing gaze behavior in multi-object scenes has yielded encouraging outcomes for detecting user intent in the context of task prediction [11]. Moreover, the misinterpretation of gaze events due to unintentional fixations or incorrect interpretation of the context caused by natural gaze behavior interferes with the correct interpretation of user intent [21]. The utilization of fiducial markers, such as AprilTags or QR codes, can improve performance by

providing the robotic system with additional information about tasks and objects, as well as assisting with robot positioning [195, 178]. The integration of a fiducial marker for each task and its subsequent linkage to an object has the potential to address the aforementioned challenges, such as robustness in selection and scalability. However, users may encounter challenges in interpreting and selecting the appropriate fiducial marker due to its visual characteristics.

In this paper, we present an approach based on pictograms to improve the robustness of daily task selection for communication between humans and robots. The pictograms' easy-to-understand appearance benefits legibility for the user. An object detection model creates interactive bounding boxes by detecting the associated task, from which the robot can obtain more information, and therefore repurposes the pictograms as fiducial markers.

The main contributions of this work are as follows:

- An object detection-based eye-tracking control strategy adapted for controlling robotic arms in daily tasks.
- A pilot dataset and object detection model trained to robustly distinguish eight common daily tasks, systematically evaluated under real-world variability.

6.2 Related Work

Pictograms are used to make environments more accessible for people with intellectual disabilities by providing alternative means of communication, as in augmentative and alternative communication. They provide rich information and enhance usability by replacing text and therefore overcoming language barriers [196, 197]. In robotics, they are used to localize the position of mobile robots [198] or to extend fiducial markers to improve comprehensibility in user-centered designs [199]. In contrast, while fiducial markers benefit robotics by providing precise object location and information about objects, the stored data is not interpretable by humans without an adequate device. Fiducial markers are used in many eye-tracking-related works revolving around task selection. These approaches use them to define the workspace and calculate the location of the desired object [13, 92, 20]. Other approaches enhance the accuracy of 3D gaze estimation via fiducial markers [68, 178]. Huang and Mutlu integrated fiducial markers to indicate desired ingredients for a task and move the robot towards the object [65]. Similar to this approach, Yang et al. and Wang et al. incorporated fiducial markers into their setup to improve detection rates and feature extraction [200, 195]. Lastly, Nunez Sardinha et al. created a system based on diegetic buttons, which realize interactive 3D-printed fiducial markers in real-world settings. Users found the system easy to learn and use [201].

Combining fiducial markers with pictograms would provide benefits from both domains: satisfactory interpretability for the user, higher information density, and enhanced accuracy in location estimation for the robot. To the best of the authors' knowledge, no approach exists that uses pictograms as fiducial markers to communicate task selection with a robot via gaze in a real-world setting.

Gaze, as an input modality, provides rich information and context for controlling robotic arms in assistive tasks, social robotics, and robotic surgery [12, 166, 25, 202]. One convenient solution for selecting tasks based on objects in a real-world scene is to interpret user intent based on natural gaze, as presented by Admoni et al., Zhang et al., and Baptista et al. [13, 22, 178]. While effective in small laboratory settings, this approach has struggled to scale to a large number of available tasks while maintaining robust selection capabilities [13, 22, 11]. Another emerging solution involves integrating vision and language, such as through the use of vision-language models (VLMs). VLMs translate images into labels and extract context from detected objects. Examples were presented by Chi et al. and Wu et al. [185, 203]. Depending on the scene and objects included, it can be challenging to conclude the correct task by identifying objects that relate to multiple tasks.

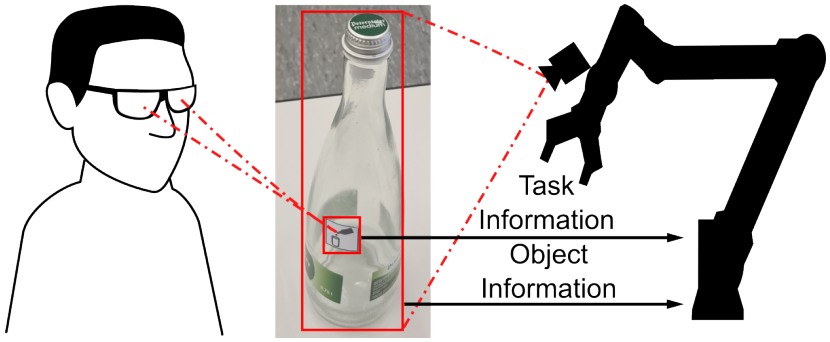


Figure 6.1: Control process: The user selects the desired task with their gaze. After a certain dwell time, the task is selected. Information is required by a camera mounted on the robot, which performs the task.

While these approaches yield practical solutions for individual tasks, it is impractical to scale them to a variety of daily tasks. Daily tasks that can be performed with a robotic arm include hobbies, household chores, and work-related tasks [162, 176], personal hygiene [169, 26], and eating and drinking independently [15, 26]. The goal of this work was to develop a reliable method for selecting between these tasks and others by using gaze information to guide the robot, while ensuring that the method can be scaled to incorporate new tasks.

6.3 Methods

6.3.1 Control strategy

The control input uses the gaze location within the eye-tracker's world camera scene. This scene includes pictograms representing the desired task, which are attached to suitable objects. Since the intended user group includes individuals with severe physical disabilities, the system design assumes that pictograms may be applied with assistance from caregivers or family members when such support is available. The user selects the desired object for the task by fixating their gaze on the corresponding pictogram. Pictograms are detected in relation to the camera scene and are selected when the user's gaze overlaps with the pictogram's bounding box. This process is outlined in Figure 6.1. The selection information sent to the robot contains the task class, object class, and object location, as well as additional information defining environmental conditions, such as the need for additional objects.

6.3.2 Pictogram design process

A subset of eight tasks was chosen based on their importance in everyday life [21]. An initial design was created for each task. Eighteen fellow researchers with mixed backgrounds in robotics and ubiquitous computing were asked to identify each pictogram's corresponding task using open-ended (free-text) responses. Based on their feedback, the designs were adjusted. Designs that appeared multiple times, such as the person, varied between pictograms to provide more distinguishing markers in the object detection training process.

The comprehensibility of the new set of pictograms was evaluated in a subsequent online survey. The aim was to investigate whether participants were able to deduce the task depicted in the pictogram. The survey was approved by the Ethics Council of the author's institution. Participants were shown each pictogram and could select one out of four task descriptions. To minimize bias, the sequence of answers was randomized in each trial. A field for optional answers was removed. They were able to provide their feedback after answering all questions. Appendix 6.A lists all selectable answers available

for each pictogram. The correct response rate was calculated to evaluate whether the pictograms are easy-to-understand. Additionally, the feedback was reviewed and summarized to gain insight into shortcomings of the pictogram design.

6.3.3 Model selection

The model was selected based on performance evaluated in the Computer Vision Model Leaderboard [204], which was benchmarked on COCO 2017. The model is expected to have a low inference time (less than 500 ms) to minimize user frustration during the control sequence by avoiding long response times of the system. The model has to have a high mean average precision (mAP50:90) at 50% and 90% for small object detection. mAP summarizes the precision and recall performance across different confidence thresholds and is used as a measure to verify if pictograms are detected robustly. Additionally, the model was chosen to achieve a high F1 score to ensure effective performance in a real-world application, where the camera position causes the size of the pictograms in the scene to be represented by only a few pixels. The YOLOv12x model [205] performed within the four best models (mAP50:95 (small) = 0.267, F1 50 (small) = 0.149, F1 75 (small) = 0.098, and general mAP50:95 = 0.527, ranking sixth).

Training performance tests revealed that YOLOv12n produced sufficiently satisfactory results, despite its lower ratings compared to the YOLOv12x model [204]. The authors decided to implement the YOLOv12n model to maintain a lightweight algorithm with low inference speed.

6.3.4 Data collection, annotation and augmentation

The revised pictograms were printed in two sizes: 1.5 by 1.5 cm and 2.5 by 2.5 cm. A total of 1,008 images were taken in domestic home and workplace settings. The pictograms were randomly placed in the scenes and partly attached to related objects. All images were captured using the same camera, resulting in a median resolution of 4032 px x 3024 px. Although using cameras with lower resolutions was considered, the idea was rejected since newer eye-tracking and robot cameras show rising performance and higher resolutions. Intentionally taking images with lower resolution to fit the current standard in data processing and object detection would result in a less suitable dataset and model for future applications. Although the number and placement of pictograms in each image were randomized, the dataset was balanced so that each pictogram class and the total number of pictograms within each image were distributed equally across samples. The number of pictograms ranged from one to five and, in some cases, represented the same task multiple times to more closely resemble real-world applications. The distance between the camera and the pictograms was randomized to reduce bias in the detected size and orientation of the pictograms.

Annotation was performed in Roboflow [206]. Due to the possibility of different levels of occlusion in both cameras, as indicated in Figure 6.1), pictograms overlapping with other pictograms or objects were annotated. Reflections were not annotated. The robot could pose a danger to the user when grasping an object due to reflections detected in glasses or cutlery, including knives, or when moving into a mirror. Pictograms that were out of focus were included to mimic real-world cases of unfocused camera images.

The augmentation was performed for the following reasons.

- Crop: 0 to 10% - Pictograms were rarely located at the edges of images. The boundaries were cropped to distribute the pictograms evenly throughout the images.

- Rotation: -15 to 15% - Simulates a higher or lower position of the camera in relation to the object and unevenly applied pictogram stickers.

- Shear: 10° - Simulates pictograms that are placed on the sides of cylindrical objects, such as cups or bottles.
- Brightness: -15 to 15% - Simulates illumination changes.
- Blur: 2% - Simulates pictograms out of focus.
- Noise 0.1% - Reducing overfit.

The total number of images in the dataset was increased from 1,008 to 2,420 with these augmentations. Three augmented images were generated for each of the 706 training images. Together with the 202 validation images and the 100 test images, the total number of images was 2,420. The images were split randomly into train (70%), valid (20%) and test (10%). The dataset is available in Roboflow[1].

6.3.5 Model training and evaluation

The model was trained with the code provided in the GitHub repository[2]. The code was partially reused from Ponomarjova et al. [207]. The dataset images were downloaded at a size of 1280x1280 px to minimize data loss due to compression. The model was trained for 300 epochs with a patience of 100.

The performance was evaluated by measuring the inference time to ensure rapid detection for the control. mAP50 and mAP50-90 were calculated to verify high overlap between the ground truth and the predicted bounding boxes so they can be implemented as interactive objects in the control architecture. In addition, precision and recall were measured to ensure the detection of all classes. The precision score was of importance to indicate whether wrong labels occur. Since bounding boxes are used as interactive buttons, a wrong prediction could lead to the incorrect selection of a task due to a wrong label. A low Recall would lead to many pictograms not detected, which would influence the functionality of the control strategy. The F1 score was calculated for completeness and comparison to the metrics provided in the initial model selection.

The model was trained and evaluated on a computer running Windows 11 with an AMD Radeon RX 6750 XT graphics card and an AMD Ryzen 5 5600X six-core processor.

The evaluation was repeated for images in which different resolutions were applied in the data download. Additionally, a dataset of 90 images was created for a stress test of the model's performance. A measuring board was used to evaluate the model performance in connection with varying environmental changes[3]. The measuring board showed 24 pictograms that had a size of 2 cm each. The pictograms were spaced out 1 cm from each other. Two different background patterns were used. With both backgrounds, an occlusion of 0%, 25%, and 50% was simulated. Taking the images with each level of occlusion was repeated 4 times, applying the occluded area to each direction once. The setting was rotated in 45° steps within a range of $\pm 90°$ to the upright position of the pictogram. No further digital augmentation was performed on the images. The dataset was annotated as stated above, but was not split into different groups due to it's purpose.

The influence of varying distances was investigated separately. All parameters remained constant, except for the distance between the camera and the measuring board, which varied from 30 cm to 200 cm in increments of 10 cm. Each object class had 16 instances. Both, the stress test dataset and these images were not uses in the training of the system.

[1] Dataset available at: `https://app.roboflow.com/wlrisemanticlables/wlri-semantic_labels/8`

[2] Model training code: `https://github.com/AnkeLinus/PictogramFidMarkerSet/blob/main/README.md`

[3] Measuring board and labeled data: `https://app.roboflow.com/wlrisemanticlables/picotgram-stress-test/2`



Figure 6.2: Upper row: Initial pictogram design, Lower row: Redesigned pictograms.

6.3.6 Initial implementation

An initial implementation was performed in which the model was integrated into a web server application that showed the camera scene and visualized the detected bounding boxes. The experiment was conducted to determine whether the inference time and precision of the bounding box location would meet the demands of the eye-tracking control. A Flask-based[4] web server was implemented to visualize camera data from an Intel RealSense D455 camera. The camera data was captured using the ROS2 Humble library provided by the manufacturer [208]. The outcome is shown in the supplementary materials video.

6.4 Results

6.4.1 Task selection

The tasks in this work were selected based on stakeholder interviews presented in previous works [21]. At least one task was selected for implementation from each category.

- Eating
- Filling a glass
- Drinking
- Switching a light switch
- Picking up an object
- Placing an object
- Brushing hair
- Scratching oneself

The tasks were chosen because they represent the challenges associated with complex tasks that include multiple objects as outlined in Section 6.3.1. They were used to explore the shortcomings and requirements of the control in the implementation step.

6.4.2 Pictogram design evaluation

After selecting the tasks, the initial design of the pictograms was created, as shown in the upper row of Figure 6.2. Eighteen fellow researchers were asked to identify the tasks represented by the pictograms to ensure understandability and collect feedback on the design. At this point, they were unaware of any answer options for the task description. The answers are summarized in Table 6.1. Most of the pictograms were interpreted correctly. Wrong answers were found in the pictograms "eat", "switch", "pick", and "scratch".

[4]https://flask.palletsprojects.com/en/stable/

Table 6.1: Overview of task interpretation of the pictograms. Researcher's CRR (Correct response rate) represents the rate the 18 researchers recognized the task from the pictogram without given answer options. Alternating answers are stated when incorrect guesses occurred. In the online study, the CRR was calculated based on 364 responses and evaluated the redesign pictograms.

	Research team		Online survey	
Task	CRR	Varying answers	CRR	Alternative answer
Eating	89%	Start of action, mouth	87%	Taking medication
Filling a glass	100%	-	89%	drinking
Drinking	100%	-	99%	-
Switch	83%	Ringing, selecting, dressing	92%	point at smth.
Picking up	83%	Release object, move up	52%	touch smth.
Placing	100%	-	97%	-
Brushing hair	100%	-	97%	-
Scratching oneself	72%	Tip of the nose, tap nose, smell	45%	Thinking, gesturing

The pictograms were redesigned for easier identification and are shown in Figure 6.2 in the lower row. The new design resolved the following issues. The robotic gripper seen in the "pick" and "place" pictogram were changed to a hand, since people without robotics knowledge have trouble understanding what it depicts. It was decided that a grasping motion depicting a hand is easier to interpret. Similarly, the spoon was moved in front of the person in the pictogram to mimic the motion toward the mouth. In the scratching pictogram, the view was zoomed out to show the person's upper body and head, and lines were added to indicate movement. This approach has been found advantageous in other publications to improve understandability for people with intellectual disabilities [209]. This increases the chance that the system will be accessible to a large group of people with varying disabilities.

The online survey aimed at investigating if the pictograms were easy-to-understand. The online survey was completed by 364 participants. In Table 6.1 the correct response rate (CRR) for each pictogram is presented. It describes how many participants have chosen the intended task description from the four available answers. Is the CRR lower than 95% the second frequently selected task is stated in the row "Alternative answers".

Participants had the opportunity to leave feedback. Of 364 participants, 58 provided written comments that contained actionable content (e.g., no emoticons or general statements indicating no issues). 25 of these responses concerned the selection options of the questions. Participants differentiated between cases in which their initial interpretation was not included among the predefined answer options and cases in which the pictogram design allowed for multiple plausible interpretations. Both response patterns had been anticipated during the questionnaire design phase. Due to the anticipated training approach for future users, the task description will be given with the pictogram to learn. Consequently, ambiguous initial interpretations are not expected to result in incorrect task selections during robot operation, although they may indicate the need for minor design refinements.

An additional answer option, giving the participants the chance to answer in free text, was not available to provoke feedback regarding the anticipated tasks, but rather to collect assumptions about what the pictogram might indicate. Therefore, the participants gave design improvements for the pictograms. The highest number of feedback reached the pictograms "scratch" with ten comments, "pick" and "place" with five comments, and "drink" and "fill glass" with four comments. These pictograms resembled the pictograms with the lowest CRR. Six answers were given regarding general design. For the learning process of future users, it was important to understand that some pictograms

Table 6.2: Model performance after training with 300 epochs. The parameters indicate performance for varying image sizes and the stress test conducted to evaluate real-world application.

Parameter	1280 px	640 px	Stress test
mAP50	0.949	0.845	0.59
mAP50-90	0.852	0.726	0.496
Precision	0.966	0.904	0.802
Recall	0.919	0.739	0.414
F1 score	0.942	0.813	0.538
Inference	420.4 ms	62.5 ms	429.2 ms

were ambiguous in their meaning until users viewed their counterparts. This was the case in "fill glass" that had as second answer option "drink" and was followed by the pictogram "drink". It was also the case with the pictograms "pick" and "place". Another design request was switching the spoon in the "eat" pictogram to a knife and fork. The use as robot control was intentionally omitted to avoid biasing participants' perception and interpretations of the pictograms. We agree that a knife and a fork would represent the task "eat" more detailed. But in future use, this could lead to unease by indicating that the robot will handle a fork and knife and move close to the user with it. The task "scratch" had worst CRR and most feedback. This resulted from the provided answer options. Many participants chose the option "thinking" instead of "scratch" and matched their feedback accordingly.

Lastly, a participant raised the question if "scratching" can be seen as task. Indeed, there is no item that handles scratching as an individual task in the IFC catalog. The classification scheme allows for interpretation that it is classified as "Looking after one‘s health", which is a overarching category of multiple tasks, that partially can be achieved with a robot. This suggests a potential discrepancy in how daily tasks are defined within assistive robotics.

6.4.3 Performance evaluation

The dataset was validated with image sizes of 1280 and 640 pixels to compare the impact of compression on performance. The dataset was evaluated using the code provided in this repository[5]. Table 6.2 shows the performance for both validations with image sizes of 1280 and 640 pixels. Due to compression, the model's performance with 640 px images was considerably worse, but required less inference. The initial implementation into the interface revealed that the pictogram size must be significantly larger relative to the camera scene to correctly detect the pictograms when the image size is set to 640 px. This procedure is demonstrated in the video provided in the supplementary materials. Pictograms were more easily detected when they had a size of 2.5 by 2.5 cm at a distance of one meter or less. Inference was sufficient to provide a pleasant control experience. To ensure functionality across the full range of a robot's workspace and detect smaller pictogram sizes, we recommend using images with 1280 px.

Figure 6.3 shows the resulting normalized confusion matrix. Each row corresponds to the ground-truth class, while each column represents the predicted class. Correct classifications are shown along the diagonal, whereas off-diagonal entries indicate misclassifications. Due to the small size of the pictograms relative to the overall image size, visual artifacts such as holes or reflections may occasionally be misinterpreted as valid pictograms, leading to 10% to 15% false positives per class. Misclassifications occurred across all classes as well as in the background category. Since the average correct detection per class is over 90% and the initial implementation showed that repositioning the camera corrects the initial incorrect detection, we consider this model sufficiently precise for developing a prototype control.

[5]Code for evaluation: `https://github.com/AnkeLinus/PictogramFidMarkerSet`

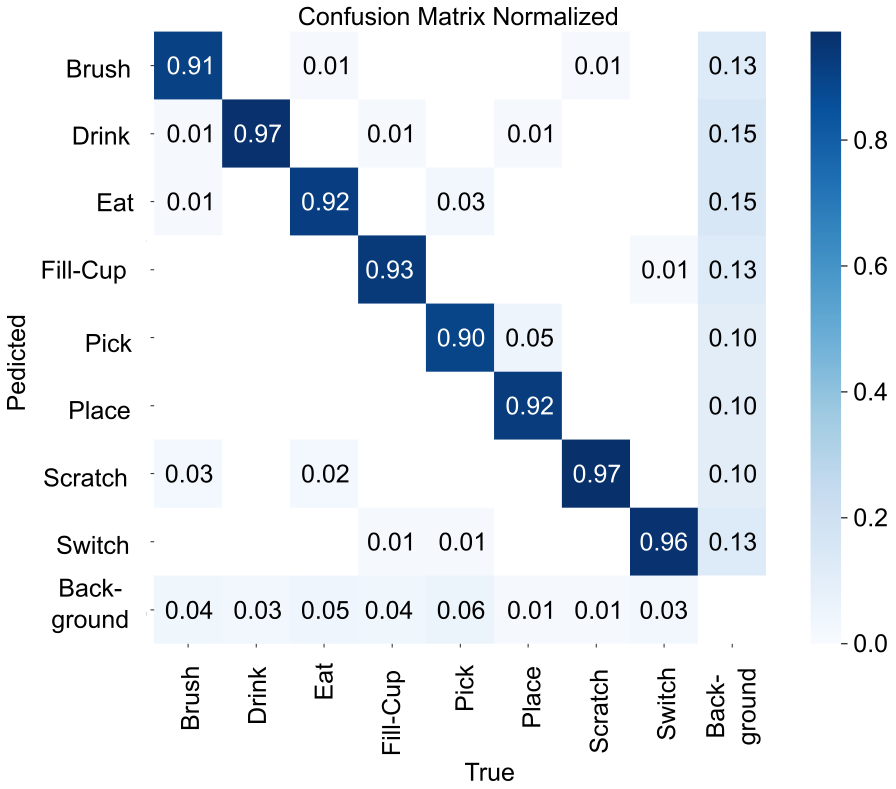


Figure 6.3: Normalized confusion matrix for test image size of 1280 px.

In addition to the classical performance evaluation, we conducted a stress test to assess the performance boundaries of the system in more detail. The background, occlusion, and rotation were systematically varied to test how the model would behave in real-world applications. In Table 6.2 the column "Stress test" shows the overall performance drop with these conditions. Occlusion was found to have the highest impact on the performance. While mAP50 was 0.958 (P=0.945, R=0.9) without any occlusion it dropped to 0.339 (P=0.346, R=0.343) with 50% occlusion. Since this drop dominated the results, the dataset was further split to investigate the impact of the background change and rotation without the influence of occlusion. The background changes also impacted the performance of the image. Precision, recall, mAP50, and mAP50-90 were lower for the patterned background condition. This effect was visible in all occlusion conditions. As an example, mAP50 was 0.985 with 0% occlusion, and a blank background dropped to 0.932 with the patterned background. With 50% occlusion, mAP50 dropped from 0.432 to 0.295 between the blank and patterned backgrounds. Rotating the images revealed no significant difference or trend in any metric. The performance varied for mAP50 in a range of 0.05.

For real-world application, these findings suggest being cautious with occluded pictograms, as it can happen in everyday life. In an ideal scenario, caregivers can help reposition the object. Regarding robot control, it is important to put objects back where they were picked up to reduce the occurrence of occluded pictograms. The performance drop in patterned environments highlights shortcomings that can be anticipated in cluttered scenes, suggesting that there should be additional tests in real-world settings.

Lastly, the impact of varying distances was tested systematically. The performance metrics are displayed in Figure 6.4. They indicate a stable performance until 70 cm distance with 2 cm pictograms, e.g., with a bounding box width and height of 85 px and a camera resolution of 4032 px x 3024 px. Between 80 cm and 120 cm (75 px to 50 px size) the model performance is visibly unstable. Recall and

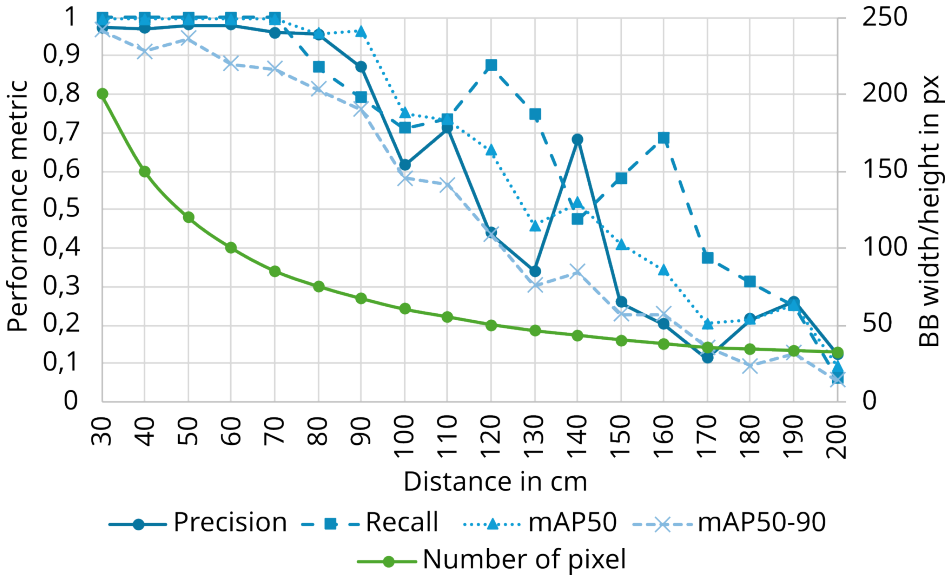


Figure 6.4: Model performance with varying distance between pictograms and camera.

precision are strongly fluctuating, and mAP50 drops significantly. The model's performance was found insufficient under these test conditions with distances greater than 130 cm. The resolution of the images is less than 46 px, which severely limits a stable detection. It has to be ensured that the detection is robust within the robot's maximum reach. For example, the Kinova Jaco arm has a maximum reach of 90 cm. Due to variations between the user's head position and the robot's base, the maximum distance between the user's eyes and the desired object including the pictogram can exceed this distance. In such cases, the unstable distance can be reached, but due to the scalable size of the pictograms, this issue can be resolved by printing them larger than 2 cm enhancing the pixel count within the pictogram.

6.5 Discussion

Object size, geometry, gaze location, and user intention influence the occurrence of unintended task selections and misinterpretations of gaze behavior in eye-tracking controls [21]. The presented approach reduces these influences by providing pictograms as visual guidance and selectable objects. Depending on the arrangement of the objects, object detection can generate overlapping bounding boxes, resulting in unintentional selection (see Figure 6.5-A). For slim objects, such as forks, object segmentation could lead to regions that are too small, causing the gaze to be localized outside the segmented object due to accuracy errors. These issues can be resolved with the presented pictogram approach, as shown in Figure 6.5-B. Since the pictograms depict specific tasks, the user can explicitly select their desired task, thereby gaining more control over the system. This, in turn, may lead to an increase in their confidence in the system's capabilities. Furthermore, because of their rectangular shape, pictograms do not display narrow object features, which makes them easier to select by gaze. Figures 6.5-C and 6.5-D show examples of the bounding box overlap in real-world scenes. Applying the pictograms creates wider gaps between the bounding boxes, reducing the error source for misinterpreted gaze fixations.

6.5.1 Reflection on key contributions

The reflection on the key contribution revolves around the question of whether the developed dataset and model meet the initial demands. In Table 6.3 the advantages and disadvantages of this approach are stated and will be discussed below in detail.



Figure 6.5: A: Overlap of bounding boxes (red) and small regions in object segmentation (blue highlighted area) are shown. B: Due to the size of the pictograms, no overlap occurs, and the region size is wider, as indicated by the black arrows. C and D: Examples of in-the-wild scenes from the dataset showcasing overlaps of object bounding boxes (red) and solved issue with pictogram bounding boxes (green).

Table 6.3: Advantages and disadvantages of key contributions.

Key contribution	Advantage	Disadvantage
Usability and accessibility	Easy pictogram application, extensible framework, customization pictogram size	Pending real-world user study, handling of complex tasks
Interpretability	Easy-to-understand, no text	Generalized task taxonomy, tasks outside the robot's ability
Scalability	Scalable to new tasks	Lack of clear object anchors
Robustness	Sufficient model performance, sufficient differentiation	Accuracy errors

Usability and accessibility

In research, reducing the workload by increasing independence is often equated with completely transferring care assistance to robots. Caregivers, family members, and potential users indicated that they prefer to use time freed from care duties for social interaction. Therefore, this approach assumes that users are in contact with a caregiver or family member for assistance with applying the pictograms and other tasks at least part of the day. This control strategy aims to facilitate the implementation of daily task approaches and generate a system capable of performing a multitude of tasks that fulfills the preferences and aims of potential users. The application of the pictograms by caregivers and family members is straightforward because they do not require alignment or focus on the scene for detection as shown by the rotation variation in the performance evaluation. Implementing these pictograms takes only a few seconds. Additionally, the selection of objects can be limited to reduce the risk of injury when handling knives or breakable objects. Larger pictograms improve detection and selection for people with ocular disabilities, like eye tremors. They can be scaled up during the printing process. The area of the virtual interactive button can therefore be scaled larger, making it easier to stay within the bounding box boundaries. Due to their quadratic shape, the overlap of bounding boxes is still strongly limited. On large pictograms with edge lengths exceeding 10 cm the desired object for interaction can be placed upon the pictogram. In this scenario, the occlusion has to be controlled to ensure stable prediction of the pictograms.

Lastly, the question arises as to whether users will perceive this approach. No studies involving people with disabilities have been conducted at this point, due to ongoing work on integrating this strategy into robot control. Considering the findings of others such as Nunez Sardinha et al. or Kudo et al., we tried to narrow down this gap. Nunez Sardinha et al. presented a similar approach based on a diegetic GUI that evaluated user perception of an eye-tracking control with 3D-printed fiducial markers [201]. Although the control strategy differs from this approach, participants quickly learned to interact with the system, which required minimal workload. Since pictograms are easier to interpret than fiducial markers, similar results can likely be found with our approach, though this must be verified in the future. Additionally, design variations proposed by Kudo et al. and shown to make pictograms easier for people with cognitive limitations to understand were applied.

Interpretability

An easy interpretation of the pictograms improves the usability by reducing frustration in learning and distinguishing the pictograms. The online study revealed that most of the pictograms are easily understandable. The initial approach of presenting the pictograms along with a description of the task

would reduce the difficulty some participants had in distinguishing between tasks. This assumption can be derived from the feedback evaluation, which shows that some participants' initial guesses were not represented among the provided answer options.

We learned two lessons from this survey. First, the survey indicated that tasks might be confused with each other. However, the limitations of what a robotic arm can perform significantly restrict the available options. This was the case in the "scratching" task in which 45% of participants selected "scratching" and 46% selected "thinking". This limits the system's scalability to the context of robotic arm use. In adjacent robot applications, the approach has to be reevaluated. The second lesson learned was that closely related tasks reduce interpretability. For example, eating was confused with taking medication and balancing an object. When integrating such tasks into the model, their pictograms must be clearly distinguishable. Additionally, a sufficiently generalized task taxonomy should be defined to avoid overly fine-grained task distinctions.

Additional improvements to the pictograms were considered to make them easier to understand but ultimately rejected. We discussed whether pictograms needed to be more unique to improve user interpretation and model detection performance. The utilization of diverse symbols in pictograms has been posited as a method to enhance detection performance. However, given humans' aptitude for pattern recognition, this proposal has been dismissed. Another improvement that was considered involved adding text to the pictograms to clarify their meaning. This idea was rejected because it is uncertain whether users are able to read and understand the presented language, which would require datasets in multiple languages. Because the pictograms are small and users can print them themselves, it is not guaranteed that the writing will be large enough to be readable.

Scalability

Some tasks lack a clear object anchor. In our set, two of the eight tasks ("scratch" and "place") cannot be directly associated with a specific object, and the task "switch" is limited applicable in public environments. While reusable pictogram stickers can be attached to objects such as plates, they cannot be placed directly on food. Therefore, this approach should primarily be understood as a high-level task selection tool for a broad set of complex activities, which can then be refined through additional gaze-based input and control modules of the community. Such controls for robotic feeding tasks are a well-researched field. Multiple authors have presented suitable solutions [210, 211, 212]. Similar approaches exist for all other tasks represented in this paper, such as drinking [213] and picking up and placing objects [58, 97]. Alternatively, intent inference and prediction methods can specify the objects and preparation steps with which the user wants to interact in a "prepare meal" task, as demonstrated by Koochaki et al. or Huang and Mutlu [14, 65]. Similar pictogram anchors must be used for other tasks, such as scratching, since the pictogram cannot be attached to the itchy area. However, the pictogram can be placed within the user's visual range and selected at any time.

Daily tasks include personal hygiene, eating, drinking, moving objects, and leisure and work activities. Often, performing daily tasks requires more than a robotic arm, such as the assistance of caregivers or additional assistive devices. In tasks related to personal hygiene, human assistance is often preferred. This is due to the potential confidentiality concerns associated with the use of robotic cameras. Petrich et al. summarize the life domains of functioning resulting from the WHODAS 2.0 test, which resulted in 25 tasks. The study identifies twelve tasks that are suitable for robotic-arm assistance and can be broken down into smaller tasks [4]. If the set of pictograms is expanded to include these twelve tasks, it is assumed that most users will be able to learn and distinguish the pictograms. Furthermore, the pictograms can be simplified to avoid overwhelming details and specify differences within these task groups, while the model's performance should not decline.

Robustness

In HRI, robustness refers to a system's ability to maintain safe, reliable, and effective interactions with humans, even when behavior, environmental conditions, or sensor input are uncertain or variable. In this context, the proposed approach promotes robustness through reliable task selection, which enables effective interaction. Intention inference could be enhanced by first selecting the desired task and then specifying the relevant time window and objects for intent inference. However, several factors in this approach can reduce robustness, including hallucinations or misclassifications by the object detection model and unintentional gaze behavior. To mitigate errors in object detection, we diversified the appearance of the pictograms and applied data augmentation strategies. The likelihood of unintentional task selection due to unintentional gaze behavior is reduced by the pictograms' small size, as shown in Figure 6.5. This modification substantially decreases the spatial region in which a gaze overlap results in a valid selection. Together, these measures improve the system's overall robustness in realistic interaction scenarios.

But as shown in the stress test, the pictogram's minimal size is limited by the resolution of the camera and the accuracy of the eye-tracker. If the pictogram is too small in relation to the image, the recall and precision will drop because the pictogram will no longer be detected reliably. Additionally, the gaze might be tracked outside the interactive button area due to accuracy errors exceeding the size of the bounding box. The maximum reach of the robot limits the distance between user and reachable objects and therefore limits to minimal pictogram size. However, not only the accuracy limits the selection process but also the overlap of the bounding box with the physical pictogram. In the stress test, mAP50-90 drop indicated that this overlap is reduced with increasing occlusion and distance. We therefore emphasize the importance of adapting pictogram size to user preferences while ensuring a minimum size of 2 cm to maintain stable detection performance and to avoid areas that are too small for robust gaze tracking.

6.6 Gaze control and robot implementation

The model and approach presented are implemented in an eye-tracking-driven control of robotic arms, as discussed in a separate work [21]. The scope of this paper is to present the control strategy in detail. The usage of the presented strategy becomes clearer when the implementation is outlined.

In this control, the tasks are selected by gazing at the pictograms as indicated in Section 6.3.1. Overlapping the gaze with the detected bounding box initiates the task execution process. Each pictogram is connected to information that defines the task, the need for a second object, and the pictogram's location in the room. A second camera mounted on the robot detects objects in the scene. Pictograms fixed to the objects were used to increase the number of features found in a feature-matching approach. This approach finds the same object in both scenes, thereby transferring knowledge about the selected task and object to the robotic system.

During a brainstorming session, other applications of this approach were discussed. It was found that it may also be transferred to other robotic applications, such as training humanoid robots, providing a quick and simple way to label objects with corresponding tasks, and facilitating communication between humans and robots during social interactions. Screens displaying GUIs for robotic arms on electric wheelchairs are not necessary with this approach since users can select their desired task in real-world scenes with eye-tracking glasses, making the setup of the system more lightweight and less power-consuming for everyday use. This approach has the potential to facilitate the integration of artificial feedback methods, as employed in augmented and mixed reality approaches [69, 178]. In these contexts, the data, when coupled with pictograms, can be utilized to articulate the selected task or visually highlight the selected task pictogram, further improving the user's understanding of the system.

6.7 Limitations

This work mentions limitations in the design process of the pictograms. Since the anticipated user group was not involved in the design process, the pictograms could be made more understandable by a larger peer group of people with disabilities. In the online survey, participants were not asked directly about physical impairments, as this information could not be verified and could potentially lead to bias in the data evaluation process.

This approach used a YOLOv12n model. Future research must evaluate whether more lightweight models or a pruning strategy can yield better results. Object detection exhibits certain variance in accuracy, as evidenced by the mAP50-90 score. If necessary, integrating fiducial markers into the pictogram itself, such as TopoTags [214, 201, 90] can enhance accuracy. Due to limited resolution related to the pictogram size, future work will investigate the potential application of such fiducial markers.

6.8 Conclusions

The aim of this work was twofold. First, a control input strategy was created to solve the ongoing challenge of correctly selecting between multiple tasks in eye-tracking-driven selection processes. Secondly, we developed an easy-to-learn, scalable, and robust control strategy for both users and robots. We presented a pictogram-based approach in which easy-to-learn pictograms serve as fiducial markers that enable the robot to locate the desired object and interpret the desired task within a scene. Eight pictograms representing daily tasks were designed. A pilot dataset containing 1,008 images from in-the-wild scenes was created and augmented to enhance its size. A YOLOv12n model, trained on this data, was used to evaluate the dataset performance and to estimate adequacy for prototyping. Precision (=0.97) and recall (=0.92) indicated sufficiently accurate detection of small pictograms in domestic home and workplace-related scenes. A stress test and online survey for evaluating the interpretation of the pictograms were conducted to show limitations of this approach in real-world application. These values indicated adequate performance for implementation into the setup, which produced promising results. Concerns regarding usability, robustness, and scalability were discussed. Key findings included the fact that scalability and robustness are influenced by the model's performance. While integrating a multitude of tasks may reduce model performance, the effect can be mitigated through varying the design of pictograms and using reasonable object anchors. Restricting the task set appropriately to subtasks that can be achieved with the robotic arm makes it feasible for the user to learn the meaning of the pictograms. We anticipate that this strategy will improve usability and task selection by defining the interaction interval and the involved objects, thereby reducing ambiguity in real-world scenes for task intent inference and explicit control inputs.

6.A Appendix

The online survey consisted of 9 questions, with questions 1 to 8 featuring a pictogram and providing answer options in the form of radio buttons. Each pictogram was introduced with the question: "Which daily activity depicts this pictogram?". Question 9 asked for the participant's feedback. In the following table, questions 1 to 8 are stated with the available answers.

Table 6.4: Overview of the answers available in each question and response distribution.

Shown Pictogram	Available answers	Number of responses
Eat	Eat	317
	Take medication	27
	Balance something	15
	Brush teeth	5
Fill a Glass	Pour a drink	326
	Drink	32
	Pour milk into coffee	4
	Water flowers	2
Drink	Drink	360
	Carry a glass	0
	Throw a can	0
	Blow bubbles	4
Switch	Operate a switch	337
	Point at something	21
	Call an elevator	3
	Show someone the way	3
Pick	Pick something up	190
	Take something out of a box	17
	Touch something	108
	Throw away trash	49
Place	Place something	354
	Build a tower out of blocks	3
	Create a route in the GPS	3
	Carry something fragile	4
Comb hair	Comb hair	355
	Massage head	1
	Put brush away	7
	Apply mascara	1
Scratch yourself	Scratch yourself	166
	Think	167
	Gesture	29
	Meditate	2

Chapter 7

Publication V: Eye-Tracking-Driven Control in Daily Task Assistance for Assistive Robotic Arms

Summary: Current eye-tracking-driven control approaches require additional sensors to estimate the exact position of an object within the eye tracker's view. The approach presented in this chapter bypasses this need, keeping the setup lightweight and eliminating the need for stationary devices to track head movements in the room. Object and task information is collected using pictograms presented in Chapter 6 within the eye tracker's field of view. A feature-matching algorithm and an object-detection model identify objects in the robot-based view and select the identical object. This technique reduces the object localization setup to an eye-in-hand configuration, enabling free head movement without the need for recurring calibration.

It is important to mention that the system allows for the continuous task selection and provides a backend for integrating robots and sensors. As mentioned in Chapter 6, the framework was designed to robustly differentiate between tasks and enable differentiation between multiple daily task groups for independent living. To this end, the framework was designed to exemplify three tasks, independent of specific robot model. It has been open-sourced to the research community for integrating individual task execution. The system's performance was evaluated through studies that simulated various scenarios that the user and robot might encounter. Although the task selection functioned as anticipated, we identified shortcomings when the user and robot faced different directions.

Auditory feedback was integrated into the system to indicate the successful fixation of a pictogram and the successful transmission of the task message, thereby reducing the workload on the visual channel. In everyday interactions, the auditory channel is used for communication and perceiving the environment. Additional auditory cues, such as beeping or text-to-speech, might increase the user's workload and lead to frustration when the system interrupts a conversation or disrupts quiet time. These considerations motivated the work presented in Chapter 8 of this thesis, which investigates the use of vibrotactile feedback.

Contribution: Anke Fischer-Janzen was the primary author of this publication. She was responsible for the conception, implementation, evaluation, and manuscript preparation. Thomas M. Wendt and Kristof Van Laerhoven contributed to the work by providing supervision, methodological discussions, and manuscript review.

Eye-Tracking-Driven Control in Daily Task Assistance for Assistive Robotic Arms

Anke Fischer-Janzen
Offenburg University of Applied Sciences
Anke.Fischer-Janzen@hs-offenburg.de

Thomas M. Wendt
Offenburg University of Applied Sciences
Thomas.Wendt@hs-offenburg.de

Kristof Van Laerhoven
University of Siegen
kvl@eit.uni-siegen.de

Abstract

Shared control improves Human-Robot Interaction by reducing the user's workload and increasing the robot's autonomy. It allows robots to perform tasks under the user's supervision. Current eye-tracking-driven approaches face several challenges. These include accuracy issues in 3D gaze estimation and difficulty interpreting gaze when differentiating between multiple tasks. We present an eye-tracking-driven control framework, aimed at enabling individuals with severe physical disabilities to perform daily tasks independently. Our system uses task pictograms as fiducial markers combined with a feature matching approach that transmits data of the selected object to accomplish necessary task related measurements with an eye-in-hand configuration. This eye-tracking control does not require knowledge of the user's position in relation to the object. The framework correctly interpreted object and task selection in up to 97.9% of measurements. Issues were found in the evaluation, that were improved and shared as lessons learned. The open-source framework can be adapted to new tasks and objects due to the integration of state-of-the-art object detection models.

7.1 Introduction

Eye-tracking-driven controls for robots and robotic arms can enable people with severe physical disabilities to regain independence in their daily lives. The ability to complete daily tasks independently has been shown to improve quality of life [3]. These controls vary in terms of their level of autonomy and interface design [12]. Various control approaches are being developed, including teleoperated systems and sophisticated methods that use gaze analysis to predict user intent and control the robot [13, 65]. When gaze is used as an explicit control input, it is often used alongside dwell-time approaches, in which the user fixates on an object with their gaze to select and interact with it. Researchers have found that this behavior comes naturally to users, as people tend to first look at the object they want to interact with in everyday situations [11].

Approaches that use fixation as an input modality rely on the accurate measurement of gaze location. Although authors often claim that eye-tracking glasses are inaccurate, their accuracy and precision have improved since the early systems described in the work of Kim et al. [39]. Interest has shifted to three-dimensional (3D) gaze estimation, which determines the point at the intersection of the gaze vector and the surface [17, 66, 19, 20]. Gaining additional depth information is beneficial because it allows for more reliable tracking of gaze on object artifacts, such as the handles of cups. Therefore, task intent can be inferred, particularly when evaluating gaze in scenes with multiple objects [14, 13]. Moreover, additional depth information can directly instruct a robot to grasp an object [10]. In many approaches, object location is estimated in relation to the user's head. Therefore, drifts in head movement measurements and slippage of the glasses can lead to significant estimation deviation.

As presented, these approaches focus on selecting and completing daily tasks. Various works demonstrate single-task implementations, such as those presented in [65, 66, 19, 20]. More recent works focus on implementing multiple tasks that can be accomplished with eye tracking and multimodal systems [185, 22]. Hagengruber et al. introduced the EDAN system, an sEMG-controlled assistive robot, and demonstrated the necessity of performing multiple daily tasks continuously and robustly

to benefit real-world users [5]. This was confirmed in findings of Farhadi et al., highlighting the importance of seamless and intuitive interaction [16]. However, most works lack this continuous task selection.

These approaches are promising in that they represent new control designs with a high level of robot autonomy. However, robust transmission of information requires high calibration accuracy in 3D gaze estimation in order to distinguish between tasks. Incorrect predictions and misinterpretations of user intent can lead to challenging situations in human-robot interaction and reduce trust in controls when the robot performs the wrong task or selects the wrong object [21]. Furthermore, selecting from a multitude of daily tasks would enhance real-world usability.

To address these challenges, we propose an eye-tracking–driven control system that reduces reliance on high gaze-tracking accuracy by shifting object localization to a robot-mounted (eye-in-hand) camera. Tasks are selected via task pictograms representing fiducial markers. Fiducial markers have been shown to enable robust object detection in unconstrained environments [90]. We transmit object information by integrating a YOLOv12n object detection model and a feature detection and matching approach, FLANN-AKAZE. Additionally, we integrated task pictograms from [215] to enable the system to distinguish between multiple tasks. We also implemented a fallback method to ensure the correct object selection. The open-source framework presented here uses only eye-tracking glasses and the robot-mounted camera for sensory input. Thus, it presents a suitable setup for controlling a wheelchair-mounted robotic arm.

Our main contributions are:

1. A comparison of feature-matching algorithms tailored to assistive robotic arm use cases used for data transmission.

2. An open-source eye-tracking–driven robot control framework based on object detection solving head movement issues by reducing robot control to an eye-in-hand configuration.

3. An evaluation of the framework that reveals key challenges and future research directions.

This work is organized as follows. In Section 7.2 works are presented covering the topics of eye-tracking-driven robot control and task accomplishment. Section 7.3 describes the anticipated human-robot interaction and provides implementation details. Section 7.4 evaluates an appropriate feature detection and feature matching approach. Section 7.5 presents the framework, and Section 7.6 its evaluation concerning the task selection success rate. In Section 7.7 the results are discussed and concluded in Section 7.8.

7.2 Related work

Recently, the development of eye-tracking-driven controls has been strongly influenced by three factors: 1) user intent prediction, 2) 3D gaze estimation, and 3) AI integration. As stated in the introduction, insights from these research fields improve the functionality and usability of using gaze as an input modality. However, this section will elaborate on the challenges that have been encountered.

Shared control mechanisms aim to balance the interaction between users and autonomous robotic systems, ensuring accurate and efficient task execution. These mechanisms enable robots to perform tasks requiring fine motor skills, improve safety by constraining robot movements, and alleviate user workload [216, 8, 165, 10]. Correct transmission of information between the eye-tracking device and the robotic system is crucial for the functionality of such controls. High error rates can cause the robot to interact with an object that the user did not select, resulting in increased frustration levels [16, 15]. The necessary sensory information is defined by the robot control strategy and the level of autonomy. The level of autonomy describes the robot's abilities and determines the user's role in the interaction

[177]. These levels range from teleoperated robots that use gaze gestures or directional gaze to fully autonomous systems that, among others, use gaze to point to the desired object [12]. In this context, eye tracking is also used to increase the data density of brain-computer interfaces (BCIs), in which gaze is used to select objects and BCI input triggers the task [64, 43].

Accurately estimating a user's gaze in three-dimensional space is essential for precise object selection and interaction in complex environments. One approach is to detect an object's location by tracking the user's gaze in a 3D scene to select the object. Several authors have implemented 3D gaze estimation in robotic applications using various techniques. For example, Cio et al. used ray tracing to determine the intersection between the camera scene and the gaze direction [19]. Li et al. calculated gaze location using gaze vectors and applied neural networks with pre-trained mapping relationships to reduce Cartesian errors [66]. Tostado et al. achieved 3D gaze tracking by developing a new calibration method that integrates the robot's trajectory to generate a continuous stream of gaze points in 3D space. They applied Gaussian processes for supervised learning [17]. Yang et al. augmented eye-tracking glasses with a camera that provides depth information. This information is used to calculate the intersection of the 2D gaze data and depth information. AprilTags were applied to enable head movement coordinate transformation, allowing users to move their heads [20]. hese approaches rely on detecting the head's position in relation to the robot to interact with objects. In this context, fiducial markers are used to define workspaces or mark objects, thereby increasing accuracy [65, 13, 92, 20, 22]. In some cases, the eye-tracking glasses are equipped with additional sensors, such as IMUs or stereo cameras, which enhance the weight of the eye tracker [217, 10] and can lead to discomfort when worn for extended periods. An eye-in-hand configuration can address this issue by providing a known kinematic chain to the camera.

Understanding and predicting user intent from gaze behavior is pivotal for making eye-tracking controls more intuitive and responsive. The natural tendency to gaze at an object when wanting to interact with it can be leveraged for this purpose [166, 11]. Belardinelli's work summarizes prediction models, including approaches using support vector machines, dynamic Bayesian networks, hidden Markov models, and others [11]. They applied this technique to mixing a drink, using fixation duration and gaze pattern sequence to predict the next ingredient [65]. Yang et al. presented an approach that uses natural gaze and applies a target-attracted gaze movement model with a Kalman filter to extract features in object- and task-related gaze patterns [200]. Gonzaléz-Díaz et al. investigated grasp prediction in multi-task learning involving tasks with multiple object interactions [158]. Predicting user intent benefits robotics by providing necessary task and goal information [13]. However, the accuracy of these predictions is limited, leading to challenges when scaling this approach to various tasks or when the user goal changes unpredictably [11, 13]. Gaze behavior, such as peripheral vision, can lead to unintended fixations [90]. Differences in user behavior complicate the training of user intent prediction models furthermore [16].

Artificial intelligence techniques play a critical role in interpreting gaze data and enabling robots to effectively accomplish tasks. Gaze can be used in multiple ways to combine AI and robotics, as demonstrated by the work of Zhang et al. [218]. Previously, approaches were presented while elaborating on user intent estimation. Others use gaze to select objects, as demonstrated by Ivorra et al. in the AIDE European project, which employed electrooculography (EOG) and electroencephalography (EEG) to estimate gaze poses. YOLOv2 (You Only Look Once) was applied to select objects in the scene [67]. Lastly, Large Language models are used to control eye-tracking robots by translating labels of selected objects into prompts for task execution [22].

Compared to the presented works, this approach differs in the following ways:

- The framework's sensory inputs are received by the wearable eye-tracking glasses as the sole input modality and one additional low-cost camera mounted to the robot.
- The user can interact with objects in the real world without needing a GUI or display.

- No 3D-estimation of the user's position is needed, which allows for free head movement and reduces the influence of calibration errors due to the eye-in-hand configuration.
- The approach uses task pictograms as fiducial markers, which allow for robust selection between multiple tasks and are easy for users to understand.

7.3 Interaction and implementation

This open-source, standalone, eye-tracking-driven robot control framework was designed to provide a foundation for integrating various robot-assisted daily tasks. To minimize accuracy errors caused by head movements, object data collection is transferred to the robot camera. This contrasts with current approaches, which use either a stationary camera or an eye-in-hand configuration. Figure 7.1 shows the general control process and computation steps that will be presented in Section 7.5. This process is presented in Section 7.3.1, which provides the rationale behind the control design choices.

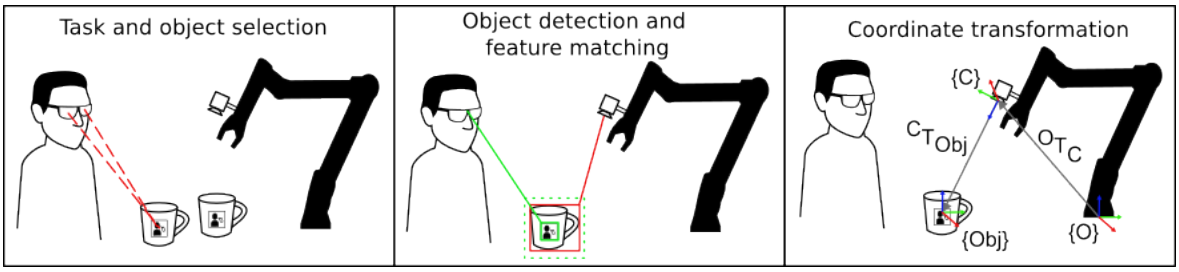


Figure 7.1: The computation process of the framework. The object is selected by gaze. The pictogram and the object are detected in both camera scenes and matched using a feature-matching approach. Coordinate transformation provides an estimated location of the object for further interaction.

7.3.1 Human-Robot Interaction process

This control focuses on articulated robots that assist people with severe physical disabilities in accomplishing daily tasks, as shown in Figure 7.1. Gaze is the only input modality. Other input modalities that depend on extremity movement or speech are not integrated. Assistive robots are primarily implemented as an extension of electric or smart wheelchairs. They are referred to as Wheelchair-Mounted Robotic Arms (WMRA), Assistive Robotic Arms (ARA), or Physically Assistive Robots (PAR) [162, 12, 174]. Due to close contact with users during activities such as assisting while eating and drinking, a predictable reaction of the robot has to be ensured. The goal of this work was to develop a reliable and robust interaction method that minimizes the need for additional peripheral devices (e.g. displays or stationary eye-tracking devices), acknowledging that people with severe physical disabilities already rely on multiple essential technologies such as speech computers or ventilators. More devices would complicate everyday usage and accessibility. Therefore, the presented framework requires only eye-tracking glasses and a RGB-D camera.

The task selection is based on the approach presented in the work of Fischer-Janzen et al. (2025) that uses pictograms as fiducial markers [215]. The pictograms are used as visual cues and represent available daily tasks that are placed on related objects. For instance, the "Drink" pictogram is attached to a cup or glass. Users initiate autonomous task performance by fixating their gaze on the corresponding task pictogram. The object attached to the pictogram is defined as an object with which the robot can interact. One benefit of this method is that new tasks can be implemented quickly by expanding the pictogram dataset and retraining the detection model. A dwell-time approach was implemented, requiring the user to gaze at the pictogram for a certain amount of time. Due to the size of the pictograms,

unintentional object selection is reduced compared to approaches that use whole objects to initiate interaction.

With this control, the robot is expected to perform the task independently. This includes trajectory and grasp planning, as well as task execution. Thus, it integrates shared control approaches. The user's role is to select the object and task and then observe and correct the robot as it performs the task.

7.3.2 Implementation details

Current eye-tracking systems attempt to minimize selection errors by improving accuracy with 3D gaze estimation methods, as discussed in Section 7.2. The presented control system addresses this issue by transferring object localization and selection to a stationary or robotic arm-mounted camera (eye-in-hand configuration). This allows for the use of heavier, more accurate cameras than wearable eye-tracker scene cameras without increasing the weight of the eye-tracker itself. This eye-in-hand configuration is beneficial for WMRA applications.

The framework was implemented in ROS2 Humble running on Ubuntu 22.04. ROS2 Humble is a Robot Operating System (ROS) framework that provides multiple libraries for robot and sensor integration. It uses standardized topics and message types that facilitate communication between modules and improve matching the framework interface to other robots. Pupil Core glasses and Pupil Capture v3.5.1 (Pupil Labs [219]) were integrated to track eye movements. An Intel RealSense D455 publishes the robot's field-of-view image data. ROS2 allows robots such as the Kinova Gen3 to be integrated using the manufacturer's repositories. Since this work focuses on general control that can be used to integrate various tasks and robots, Section 7.5.3 will discuss the implementation of robots and tasks.

7.4 Method exploration

Transmitting data from eye-tracker to robot camera is done with feature matching. A comparison between algorithm was performed to find the best solution for the control. In this section we present the algorithm analysis.

The OpenCV library offers established feature-detector-descriptors (FDD) and feature matching (FM) algorithms. However, the available options were limited by the version of OpenCV (version 4.5.4) used by other modules to match image variables between object detection models and ROS2 messages. The following feature detectors and descriptors are available: Accelerated-KAZE (AKAZE) [220], KAZE [221], Binary Robust Invariant Scalable Keypoints (BRISK) [222], Oriented FAST and Rotated BRIEF (ORB) [223], Scale-Invariant Feature Transform (SIFT) [224], Speeded Up Robust Features (SURF) [225], Adaptive and Generic Accelerated Segment Test (AGAST) [226], and Maximally Stable Extremal Regions (MSER) [227]. In the following, the feature matching algorithms Brute-Force Matcher (BF) and FLANN-based Matcher (FLANN) are evaluated. The algorithm implementation used to compare each algorithm was based on the tutorial provided by OpenCV [228].

The evaluation was based on the idea that the object selected in the eye-tracker scene is the region with the greatest number of matched features in the robot scene. The bounding box generated in the robot scene that is closest to this feature cluster is most likely to represent the desired object. After the initial testing presented in Section 7.4.2, the approach was refined to improve computation time and accuracy. This refinement involved comparing the bounding boxes of the eye-tracker camera scene with the bounding boxes of the robot camera scene. The number of features that are matched between two bounding boxes indicates the degree of object matching.

Table 7.1: A comparison of publications on FDDs and FM algorithms based on OpenCV.

Source	OpenCV Version	Dataset	Recommendation
Tareen and Saleem (2018)[229]	OpenCV 3.3, Matlab 2017a	Images from University of OXFORD, MATLAB, VLFeat, OpenCV	SIFT, BRISK for highest accuracy and ORB, BRISK for highest efficency.
Noble (2016) [230]	OpenCV 3.1, C++	Benchmark image set "sculpture"	BRISK-BF for highest efficiency, SIFT-FLANN for most features matched.
Pusztai and Hajder (2016) [231]	OpenCV 3.0	Self-made Dataset	SURF, KAZE, AKAZE depending on images for highest accuracy.
Golovnin and Rybnikov (2021) [232]	OpenCV 4.5.1 Python	Lund University's motion pipeline data set	ORB-FLANN

7.4.1 Comparison with literature

A literature search was performed that included four well-established or recent publications covering the comparison of OpenCV based feature detection and matching algorithms [229, 230, 231, 232]. Based on the insights gained from these publications, differences between versions and comparison methods were identified. Algorithms were excluded from further testing based on the authors' recommendations. Details are stated in Table 7.1.

Some of the sources are ten years old, which limits their comparability with newer versions of OpenCV. This is especially true since different versions of OpenCV use different underlying programming languages. The available algorithms varied between versions. For instance, the SIFT patent expired in 2020, so it was excluded from older studies. Comparability was limited within some works due to the small number of detected features in ORB and GFTT, which were capped at 500 and 1,000, respectively. [232, 231]. Therefore, comparing performance, especially for feature detection and matching time, might be biased. The works of Golovnin and Rybnikov [232] as well as Pusztai and Hajder [231] found that FLANN-based FM performed worse than brute-force matching. Furthermore, both sources indicated that combining feature detectors and FMs can drastically affect performance, suggesting that feature detectors must be selected based on the FM.

The varying results led to the decision to repeat the tests using custom test cases designed for the robot application. This will allow to evaluate the performance of the FDDs and FM algorithms.

7.4.2 In-the-wild search approach

Two test cases were recorded and referred to as FM test cases. Both cases illustrate real-world scenarios, as demonstrated in Figure 7.2. The first test case depicts a set table in a home setting. The second case includes a table set up outside on a sunny day. These differences in lighting conditions and backgrounds were chosen to gain more detailed insight into the performance of the algorithms. From the field of view of the eye-tracker, bounding box cutouts are generated for the objects whose features will be used as descriptors in the FM algorithm. The image below simulates the field of view of a camera mounted on a robotic arm attached to an electric wheelchair. Most eye-in-hand configurations describe attaching the camera to the flange behind the end-effector. Therefore, cameras in this configuration are in an elevated position, closer to the scene than the user. In the indoor scene, the plate is partially outside the field of view, which matches real-world scenarios due to the natural misalignment between the user's view and the robot camera's view. In the second robot view, the camera is positioned on the

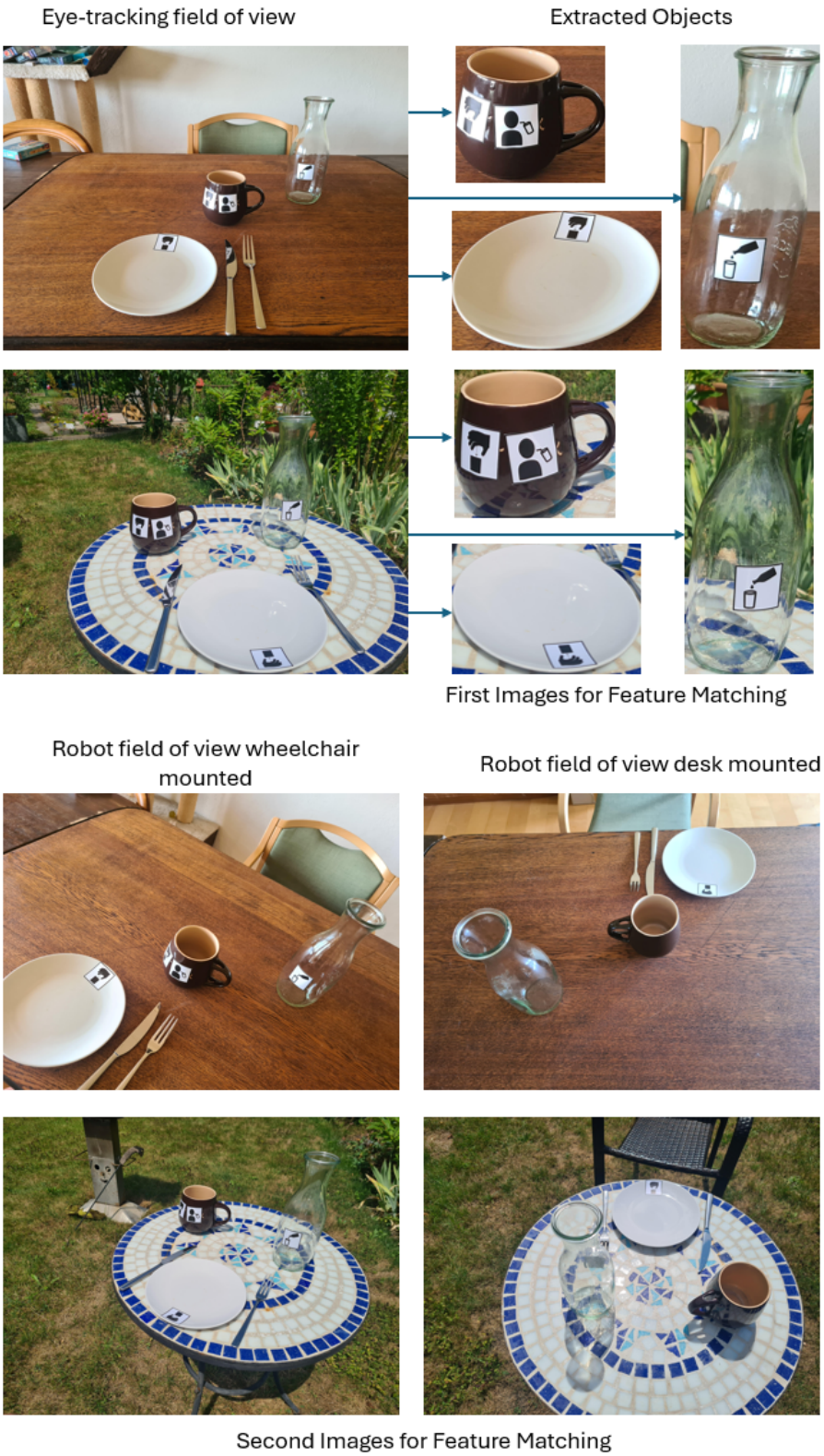


Figure 7.2: FM test cases. Upper scene: Inside scene of a set table. Lower scene: Outside scene of a set table in direct sunlight.

Table 7.2: Algorithm comparison results. All values represent the mean value calculated from each match per algorithm.

Algorithm	Total computation time (sec), mean (± SD)	Computation time/feature matched (ms), mean (± SD)	Features matched, mean (± SD)
BF-AKAZE	1,728 (±1.239)	1,371 (±0.886)	22,25 (±22.52)
BF-BRISK	3,795 (±2.296)	3,776 (±2.286)	10,83 (±9.81)
BF-ORB	0,407 (±0.455)	0,326 (±0.108)	12,5 (±10.33)
BF-SIFT	12,40 (±3.109)	12,243 (±2.997)	23,41 (±12.52)
FLANN-AKAZE	0,551 (±0.477)	0,194 (±0.124)	22,83 (±23.26)
FLANN-BRISK	0,537 (±0.349)	0,517 (±0.339)	11,83 (±9.91)
FLANN-ORB	0,377 (±0.356)	0,003 (±0.009)	17 (±16.75)
FLANN-SIFT	1,229 (±0.548)	1,068 (±0.437)	35,08 (±15.86)

robotic arm on the table across from the user. This configuration could be used for feeding tasks or to maximize separation of the shared workspace. Compared to standard feature-matching applications such as merging single photos into panoramas or video stabilization, this setup introduces challenges such as feature occlusion and misaligned edges. However, there are applications, such as feeding robots placed on the table, that require this setup.

The OpenCV 4.5.4 2D Feature Framework contains the following feature detection and description classes: ASIFT, AGAST, AKAZE, BRISK, FAST, GFTT, KAZE, MSER, ORB, and SIFT, and descriptor matchers: Brute Force Matcher and FLANN-Based Matcher. The excluded algorithms were ASIFT, AGAST, FAST, and GFTT, as they function as a keypoint detector with a missing descriptor and are implemented in other algorithms, such as ORB. KAZE was excluded because AKAZE demonstrated superior performance in the presented literature. MSER was excluded because it has no sophisticated matcher [231]. The included algorithms were therefore AKAZE, BRISK, ORB, and SIFT. Both FMs were used in the experiment and paired with all FDDs, resulting in eight algorithm pairs.

Each permutation of FDD and FM was tested with the FM test cases. This resulted in ten feature-detection-description instances for each individual image (cup, plate, bottle, robot on wheelchair, and robot on a desk in each inside and outside scene). The best algorithm was selected based on computation time per feature matched, total computation time, and number of matched features after applying a distance ratio test suggested in the tutorial of OpenCV [228] and introduced by Lowe [224]. The mean and standard deviation (SD) of the parameters were calculated for each algorithm pair and are shown in Table 7.2.

The results from this test and literature were compared. It was confirmed that the fast computation time of ORB is due to the reduced number of 500 detected features, as stated in [231, 232]. Regarding the computation time for each feature detected, BRISK (0.019 ms/feature), SIFT (0,16 ms/feature) and AKAZE (0,35 ms/feature) were faster than ORB (0.37 ms/feature) as standalone FDDs. As highlighted in Table 7.2, both BF-ORB and FLANN-ORB were the fastest algorithms in terms of total computation time, showing the stated computation time difference between FDDs and FDD-FM-pairs, and confirming the results of Golovnin and Rybnikov. This is due to the low computation time per matched feature in BF-ORB, which outperforms the other algorithms. SIFT with both FMs had the highest number of matches, followed by FLANN-AKAZE, BF-AKAZE, and FLANN-ORB. As indicated in Table 7.1, results of Noble et al. were confirmed for SIFT-FLANN, yet BRISK-BF is now one of the slowest algorithms.

Based on these initial results the performance of FLANN-SIFT due to the high number of features matched; BF-ORB, due to its speed; and FLANN-AKAZE, due to good performances in both parameters



Figure 7.3: Feature matching results for selected algorithms. In the best cases, shown on the right 32 (BF-ORB), 87 (FLANN-AKAZE), and 53 (FLANN-SIFT) features were matched and not removed by the ratio test. In the worst cases, shown on the left 2 (BF-ORB), 3 (FLANN-AKAZE), and 12 (FLANN-SIFT) features were matched. Color changes in contrast to Figure 7.2 result from using OpenCV library and have no impact on performance.

were further investigated. The matched features were visualized for each algorithm as illustrated in Figure 7.3. It is evident in FLANN-AKAZE and FLANN-ORB that both algorithms detect the edges of the pictograms well. However, FLANN-SIFT detects and matches blobs, resulting in more diffuse matching. In the scenario with the robot mounted on the desk, the pictograms are occluded, resulting in worse prediction for all algorithms. FLANN-SIFT had the highest number of outliers in the best and worst case.

In conclusion, the in-the-wild search approach uses FLANN-AKAZE due to the trade-off between the large number of matched features and total computation time, as well as the relatively small number of outliers.

7.4.3 Comparative approach

As presented, the accuracy of correctly matched features can vary drastically with robot position. The comparative approach was chosen to exclude areas that are not detected as objects, such as the tables in Figure 7.3. Matching the bounding box cutouts from both camera views was expected to yield the highest feature-match score for the pair depicting the same object. Due to the reduced image size searched for features, this method greatly reduces the computation time needed for algorithms such as SIFT.

A pairwise comparison of this approach was performed with BF-ORB, FLANN-AKAZE, and FLANN-SIFT due to their previous performance. The 12 instances observed included matching features in the user's view cutouts of a cup, bottle, and plate in scenes combined with the objects cutouts of the scenes indoor-desk, indoor-wheelchair-mounted, outdoor-desk, and outdoor-wheelchair-mounted. The ratio was set to different levels 0.6, 0.65, and 0.75 to estimate the occurrence of false positives and number of matched features. In the case of 0.6 and 0.65 ratio ORB did not find any matches in 3 of 12 instances, making it unusable for this approach. Most features were removed from the ratio test due to a high number of false matches for SIFT, with all ratios. Based on this outcome, FLANN-AKAZE was implemented in this approach. FLANN-AKAZE predicted the object correctly in 11 out of 12 instances, showing the highest number of matches. Changes in the ratio did not influence performance. In the outside scenario, it correctly matched objects more often than the other algorithms. Since it did not exhibit issues with zero values in feature matching, as ORB did, it was preferred. It was concluded that the algorithm is more robust in varying scenes and requires less fine-tuning.

7.5 Framework description

In this section, we present the overall system framework, outlining the data flow from user interaction to robotic execution, that was introduced in Figure 7.1. Accordingly, in Section 7.5.1, the communication with the eye-tracking device and pictogram detection used to realize a real-world gaze cursor are presented needed to initially select the object. In Section 7.5.2 the transfer of data from eye-tracker to robot camera is discussed. The fallback mechanism to minimize the framework's arbitrary behavior and error sources incorporating both the in-the-wild search approach and the comparative approach is presented. This mechanism acknowledges the occurrence of misclassified bounding boxes and provides a reliable way to address such situations. Once the data has been successfully transmitted, the framework conceptually transitions to object interaction. Although no physical robot execution is included in this work, Section 7.5.3 describes the essential principles for estimating the object's pose relative to a robot as a prerequisite for interaction.

7.5.1 Eye-Tracking communication and task selection

The open-source framework is available in this link[1]. The EyeTrackerCommunicator, TaskDetection, and GazeCursor modules are described in this section. Communication with the eye-tracker was established using the Pupil Core Capture app. This app was further used for calibration, which takes less than a minute and is the only hardware that needs calibration before using the framework. The API backbone provided by Pupil Labs was integrated to ensure communication. In the EyeTrackerCommunicator, the "frame.world" and "fixation" topics were used to retrieve the necessary data published by two ROS2 Humble publishers created by the authors. The world video is published as an image message, and the fixation data is given as normalized x and y coordinates, with the lower left corner of the world video at 0/0 coordinates.

[1]Link to repository: `https://github.com/AnkeLinus/EyeRAC`

The GazeCursor logic employs a dwell-time approach, in which the user must maintain visual fixation on the object for a defined duration to confirm the selection. The API backbone automatically publishes an occurring gaze fixation, with a dwell time depending on the dispersion and duration of the detected fixation within the field of view [233]. With pupil detection confidence greater than 0.95 at all times, the dwell time, and therefore the update rate of the publisher, is less than one second. Pictograms are detected and published in the TaskDetection module. The integrated model's training and augmentation used for pictogram detection is described in [215]. This model is trained on eight pictograms for the tasks "Drink", "Fill Cup", "Eat", "Scratch", "Switch Light Switch", "Brush", "Pick Object", and "Place Object".

The algorithm tracks new fixations and pictogram bounding box locations in relation to the eye-tracker scene camera frame. Overlap is calculated based on the locations. As soon as an overlap between a bounding box location and a fixation is found, the corresponding pictogram is selected. Then, a message is sent to the FeatureMatching node of type "Features". This message includes the selected bounding box information, such as location, width, height, and class, as well as task information and information about the existence of a secondary object.

Reviewing the available tasks, it was evident that some require a secondary object to be completed. For example, the "Fill Cup" task requires at least one beverage and one cup or glass. The framework was enhanced with a secondary object selection. In these cases, the task is selected as usual, and then the user can select an object with or without an attached pictogram. This feature could be useful in social situations, such as a shared meal, allowing the user to select and fill guests' glasses.

7.5.2 Human-robot information transfer

The user's selected object and the robot's interpretation has to be matched with high accuracy to ensure correct selection. To achieve this, the cutout of the pictogram is scaled up, so that the object is completely visible. Then both bounding box cutouts from the eye-tracker scene and from the robot camera scene are compared with FM. As seen the resuming falsly matched features still can lead to issues in object data transmission. Therefore, the aforementioned approaches were combined in a fallback mechanism to improve the selection process. The fallback mechanism is presented in Figure 7.4 and checks if the bounding box class IDs match in both images. If this is true, there is a higher chance that object detection already indicates the selected object's existence. Hence, the comparative approach is chosen. Past experiments have indicated that, due to low representation in the training set, some classes may lack sufficient detection. For example, an object of the "cup" class may be detected as the "suitcase" class. In this case, the fallback mechanism clusters the features found in the complete robot camera image and calculates the distance to the closest bounding box. This behavior was realized with the in-the-wild search approach. The task and object IDs were matched using a custom library that converted the task class IDs into the object class IDs available in the MS COCO dataset and the assistive dataset [207].

Implementation of comparison approach: The bounding box cutout, representing the object attached to the selected pictogram is used to match detected features in all found bounding box cutouts. The cutout with the highest number of matching features will indicate the selected object for the robot. It is tested again to ensure that the object class of the final selected bounding box is identical to the necessary object for completing the selected task. As presented in the feature matching algorithm selection, the matching rate can be extremely low when the robot is mounted on the opposite side of the user (desk test case). Comparing the bounding box class can improve detection robustness and prevent incorrect predictions.

Implementation of in-the-wild search approach: If the class is not found in the robot view cutout set, then the robot field of view is used for feature matching. Clustering is then applied to identify the area with the highest density of matched features. The number of clusters is determined by the number

of objects found in the scene, plus one additional cluster to account for outliers and missed bounding boxes. In this approach, k-means clustering was used to cluster matched features locations in the image of the robot field of view. The functionality of both approaches is evaluated in Section 7.6.

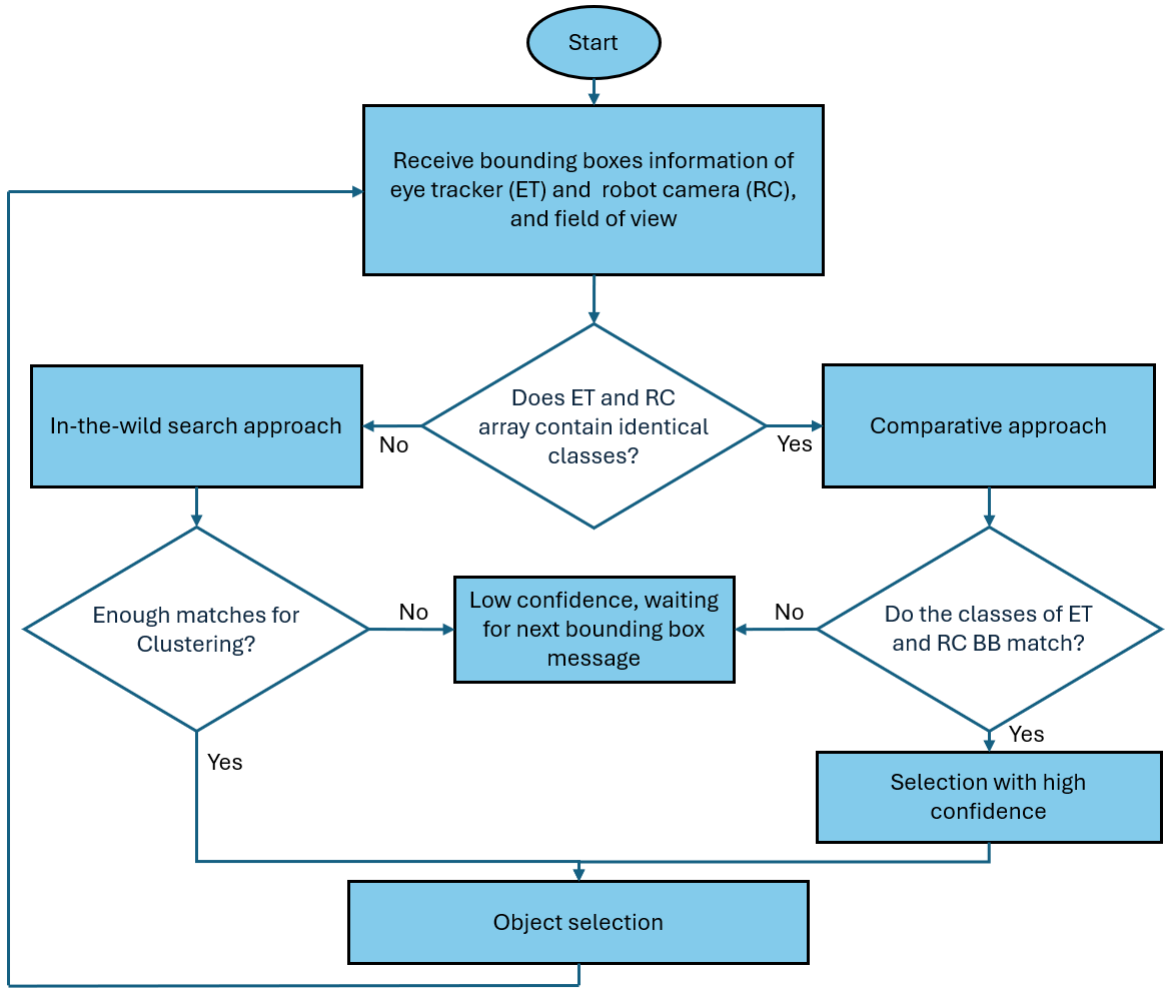


Figure 7.4: Program flow for robust selection information transmission between eye-tracker and robot field of view.

7.5.3 Shared control interface

Due to the reduction to an eye-in-hand configuration, coordinate transformation can be used to estimate the location of the object. Therefore, the accuracy is only limited by the errors of the robot, camera, and setup assembly. The coordinate transformation is illustrated in Figure 7.1 and can be described as stated in Equation 7.1, with the world coordinate frame O, the camera coordinate frame C and the object coordinate frame Obj. p describes the position of the corresponding coordinate frame in relation to the reference coordinate system.

$$^{O}p_{Obj} = {}^{O}T_{C} \cdot {}^{C}p_{Obj} \tag{7.1}$$

With ${}^{O}T_{C}$ describing the transformation from the world to the camera coordinate frame.

$${}^{O}T_{C} = \begin{bmatrix} {}^{O}R_{C} & {}^{O}p_{C} \\ 0 \quad 0 \quad 0 & 1 \end{bmatrix} \tag{7.2}$$

One suitable approach to obtaining model information from an object is to transform a 2D image cutout based on the bounding box into a 3D point cloud captured by the camera. However, since the Intel RealSense camera provides separate depth and color streams with different resolutions and principal points, depth pixels cannot be directly aligned with color pixels by index. Instead, depth values are reprojected into 3D using the depth intrinsics (Equation 7.3), then transformed into the color camera reference frame through known extrinsics (Equation 7.4), and finally reprojected into the color image using its own intrinsics (Equation 7.5). A 3D point from the depth map is associated with the object when its corresponding color-image projection satisfies the bounding box inclusion condition defined in Equation 7.6.

$$\begin{bmatrix} X_d \\ Y_d \\ Z_d \end{bmatrix} = Z_d(u_d, v_d)\, K_d^{-1} \begin{bmatrix} u_d \\ v_d \\ 1 \end{bmatrix} \tag{7.3}$$

$$\mathbf{P}_c = R_{dc}\,\mathbf{P}_d + \mathbf{t}_{dc} \tag{7.4}$$

$$\begin{bmatrix} u_c \\ v_c \\ 1 \end{bmatrix} = K_c\, \frac{1}{Z_c} \begin{bmatrix} X_c \\ Y_c \\ Z_c \end{bmatrix} \tag{7.5}$$

$$u_{\min} \leq u_c \leq u_{\max}, \quad v_{\min} \leq v_c \leq v_{\max} \tag{7.6}$$

Libraries such as GPD (Grasp Pose Detection) [234] can generate grasps based on an object's point cloud. Reducing the number of point cloud voxels representing the background can enhance the quality of the generated grasp. Preliminary tests showed that object segmentation methods produced adequate results. The framework's structure allows for the implementation of other tasks as well. The approaches are presented in Section 7.7.2.

The resulting framework, including the introduced modules, is visualized in Figure 7.5 to facilitate usability for other researchers.

7.6 Evaluation

The evaluation focused on the reliability of object selection and data transfer. Specifically, we report a task selection success rate indicating how often the pictogram-associated object was correctly communicated to the robot side of the framework. A user-centered evaluation was not included, as comprehensive user satisfaction assessments inherently depend on additional components such as robot behavior, interaction dynamics, response timing and safety measures, which are not implemented within the scope of this work. Therefore, a meaningful usability study would require the fully integrated shared control system, which is reserved for future work.

7.6.1 Test environment

The test environment was setup under laboratory conditions to ensure comparability and reproducibility of the data and cases. Auditory cues were added, one when a pictogram was successfully selected and one when the according message was published, to inform the users that they can proceed. This greatly facilitated interaction in the experiment. One participant was recruited from the authors team to conduct the experiment. The participant was asked to wear the eye-tracking glasses and calibration was performed. The participant was instructed to gaze at the pictogram until the auditive cue was heard, then looking briefly at a random location away from the pictogram, followed by looking at the next pictogram in a defined sequence. After activating the robot camera via the ROS2 launch file and

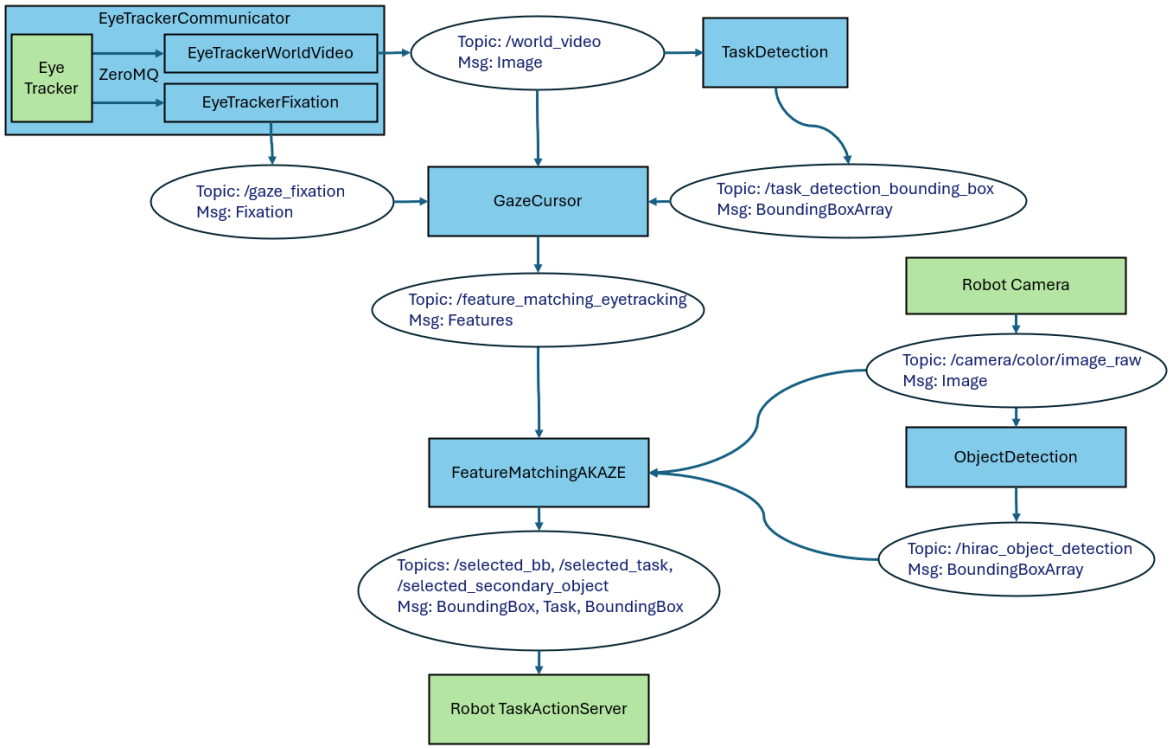


Figure 7.5: The framework of the eye-tracking controller. The depiction shows communication between nodes based on the ROS2 RQT graph. The rectangles represent nodes and the circles represent topics with message types. The green rectangles indicate interfaces to the periphery.

starting the framework, the measurements began. Three cases were tested, referred to as SU cases (SetUp cases) to distinguish them from the test cases in the feature matching evaluation. Before each test case, the eye-tracker was calibrated. These three cases examined how the framework reacted to the following aspects.

- **SU Case 1:** Is the object and task information transmitted correctly to the robot framework?
- **SU Case 2:** How are situations with falsely labeled or detected tasks and objects handled?
- **SU Case 3:** Can the framework detect a varying scene between user and robot field of view?

In all SU cases, the interactable objects were a cup, a fork, and a bottle as seen in Figure 7.6. The combination and tasks attached varied between cases as stated below.

SU Case 1: This case was used to represent perfect object detection and labeling. The cup, bottle, and fork in the upper row of Figure 7.6 were placed in front of the slightly elevated robot camera. Correct object detection of all objects was ensured at any point in time by placing them at distinct points in the scene. The cup had attached the pictogram for task "Drink", the fork the task "Place", and the bottle the task "Pick". The user sat next to the robot camera, wearing the eye-tracking device, in approximately 80 cm to 100 cm distance to the objects. The comparison approach was tested with this setup.

SU Case 2: This condition was created to check how the system handles detected objects with wrong label. All task pictograms on the objects in Case 1 were replaced with the task "Place". Additionally, the class of the object in the task-to-object library was switched to a non-existent object class to ensure that in no case the object could be matched. This has lead always to the in-the-wild search approach. The participant took the same position and behaved similarly to SU Case 1.

SU Case 3: False positives were evaluated. In this scenario, it was assumed that user and robot do not face the same scene, such as in the case of the user facing away from the robot to talk with someone or just observing the surroundings. In this case, all objects illustrated in Figure 7.6 are used. The upper row of objects in the figure is placed in the robot's field of view, while the lower row is placed outside the robot's field of view and inside the user's. The SU case was conducted twice. First, all pictograms were attached, as shown in the figure. Second, the pictograms were removed from the robot scene to evaluate their influence on FM. The participant was seated in the same room but facing away from the robot's field of view to ensure they observed different scenes.



Figure 7.6: Test objects and attached pictograms. The upper row with white cup was used in all SU test cases. The lower row with brown cup was used for SU test case 3.1 and 3.2.

7.6.2 Test parameters and data preparation

The ratio was set to 0.75 for the comparison and in-the-wild search approach. The task selection success rate and mean task selection time were calculated. Since the selection process sometimes triggered multiple times due to their fast reaction time, duplicates were excluded from the measurements. In the following, the number of included measurements will be stated. In each SU case at least 100 measurements were recorded, each describing one task selection via gaze. Since the measurements and repetitions were counted by the participant, the number of measurements may vary from 100 due to human error. Additional data was recorded, such as time stamps and resulting features left after the ratio test, as well as cluster sizes. This data was used to look into error sources.

7.6.3 Evaluation Results

SU Case 1: The eye-tracking glasses had an accuracy of 0.545° and a precision of 0.096°. In total, 156 measurements were recorded, of which 74 were included in the evaluation. A successful message was only sent to the robotic framework when the feature matching could determine the object with the maximum features matched in the comparison approach, and if the object classes matched for the selected object between user and robot scene. These limitations led to a task selection succes rate of 97.1%, in which the data of the object and task selected by the user was correctly transmitted to the robot. The task selection time, from the selection of the object until the transmission of the message, was 260.3 ms (SD 88.5 ms).

SU Case 2: In this case, the objects' classes in the robot scene are not matching with the classes predicted from the eye-tracker. For example, task "Drink" assumes the existence of a cup, but "cup" classes are not present in the bounding box array of the robot scene. The eye-tracker had an accuracy error of 0.629° and a precision error of 0.072°. 186 measurements were recorded. Excluding duplicates, 102 measurements were included in the evaluation. In 59 measurements (57.8%), the message was sent, of which 35 messages predicted the bottle correctly, 23 predicted the fork correctly, and 1 message predicted the cup correctly. An assumption why information for cup was so sparsely sent might result in few matched features by FLANN-based matching. The ratio test removed all matches found in comparison to the other classes, even though the amount of matches found in all classes were similar, ranging between ten and twenty. Samples for all classes revealed that the locations of matched features for the cup scattered more in the image than for the other two objects. Since all pictograms are the same, the features matched with the pictograms are ambiguous. The features matched in the rest of the image are the deciding factor. The white cup on the white table might be confused with the transparency and reflections on the bottle or fork. Therefore, the bounding box classes mismatch and the message will not be sent. In the other 42.2% of measurements no message was sent due to an insufficient number of matches. Yet, it is rather better to not send a message than a wrong message. This could lead to a unpredicted task completion, that might lead to user dissatisfaction and lost sense of agency.

SU Case 3: The first trial with attached pictograms in both camera scenes, 168 measurements were recorded with the eye-tracker accuracy of 0.629° and precision of 0.072°. 95 measurements were included after removing duplicants in the evaluation, of which 45 messages were sent (47.4%). This indicates that these measurements would result in a false positive and selection of the wrong object due to the separated field of view of user and robot.

It was assumed that removing the pictograms from the robot set would reduce the false positive rate. The test was repeated, leading to 98 included measurements, of which 61 measurements were sent (57.4%). The higher number of sent messages could result from varying lighting conditions influencing the feature matching, since the room was lit by a window and the measurements were not done at the same time of day. We will discuss the effects between SU test cases and the effects on interaction in Section 7.7.1.

7.7 Discussion

In the following, the found effects are discussed. These insights were used to explain possible optimization strategies to fine-tune the control. Lastly, the integration of task routines for robots are discussed.

7.7.1 Lessons learned

In summary, the method exploration revealed that AKAZE showed the trade-off of fewer false positives and faster computation time. Consequently, the method was used in the comparison and in-the-wild approaches. Performance was tested in the framework evaluation. As stated in Table 7.3, the FLANN-AKAZE algorithm achieved adequate results. The performance of the eye-tracking-driven control was evaluated in different cases that investigated potential challenges of such a control. In Table 7.3 the SU case results are summarized.

Overall, the number of messages sent was low in all categories, ranging from 44.9% to 57.8%. Messages were not sent when objects were classified differently than expected in the eye-tracking task selection. For example, the "drink" task was selected, but a fork was identified as the most matching object in the robot camera scene. Differences between the approaches were found indicated by the task sent rate. SU cases 1 and 3.1 used the comparative approach, while SU case 2 used the in-the-wild search approach. SU case 3.2 also used the in-the-wild search approach in 31 out of 61 cases. As

Table 7.3: Summary of the evaluation test cases, task selection success rates, and correct messages sent. Task sent rate refers to all measurements, while task selection success rate and messages sent for objects are calculated on the basis of the total number of messages sent.

Test case	Total measurements	Messages send	Task send rate	Task selection success rate	Messages correctly send for bottle	cup	fork
SU Case 1	74	34	47.3%	97.1%	26	4	4
SU Case 2	102	59	57.8%	92.8%	35	1	23
SU Case 3.1	95	45	47.4%	97.9%	9	10	26
SU Case 3.2	98	61	57.4%	94.9%	40	21	0

can be seen, the task sent rate is lower with the comparative approach, indicating stricter behavior that improves the task selection success rate. The low number of messages was considered a minor drawback, as the average task selection rate was 260.3 ms per calculation. Since other studies have shown that acceptable dwell times can be twice this duration [31], it is assumed that the waiting time for a repeated measurement is appropriate for users.

With a mean task selection success rate of 95.67%, incorrect predictions about the transmission of the object are rare. However, as tested in SU Case 3, this does not always result in good selection rates. Due to the experimental design, it was expected that the task sent rate would drop near zero in cases 3.1 and 3.2. There were fewer matched features compared to the other test cases. Since the minimum threshold for matched features was set too low, the message was sent anyway. The threshold was implemented more strict in the code, resembling one lessen learned.

By comparing the messages sent for each object, we deduced that the following aspects could influence the selection: Due to the test setup, it was assumed that each object was selected equally since participants had to follow a sequence when looking at the objects. Duplicates were removed before calculating the number of messages sent for each object class. In each case, one of the objects was correctly detected more often than the others. In SU cases 1, 2, and 3.2, the bottle was detected more often. As soon as the in-the-wild search approach is conducted, the bottle has a slight advantage of being selected since it has the biggest bounding box and covers the most space in the camera scene, as seen in SU cases 2 and 3.2. SU case 1 used the comparison approach. Due to the size of the bottle's bounding box, a larger cutout was created, allowing AKAZE to detect more features. This contrasts with SU case 3.1, in which the fork is the most correctly detected object and the pictogram leads to a high number of matched features. In SU case 3.1, where the robot and participant viewed scenes with varying objects, the fork was selected most of the time. One reason might be that the selected forks did not vary as much as the bottle and cup. Unlike in case 3.2, the pictograms were applied to both sets of objects, which may have increased the number of matched features.

As indicated, the potential advantages of pictograms in this feature matching approach compared to other selection methods were examined. In conclusion, it is evident that pictograms impact the feature matching algorithm. When comparing Case 3.1 (where the objects had identical pictograms) and Case 3.2 (where the pictograms were removed from one set of objects), the task selection rate was not significantly different; however, smaller objects, such as forks, were easier to detect. Facilitated feature detection with pictograms was also found when comparing the two robot positions in Figure 7.3. Pictograms have edges and varying designs with high contrast that represent the features searched for by AKAZE. Different results may be obtained using other FDD algorithms.

The lessons learned from the evaluation can be summarized as the following three approaches that were designed and integrated into the control. Before using this framework researchers should apply these lessons to fit the framework to their use case.

1. Calibrating the resizing of bounding box cutout sizes in robot scene
2. Calibrating the threshold for minimum number of matched features
3. Choosing appropriate ratio for ratio test

The threshold and ratio were integrated as parameters into the FeatureMatching module and can be adjusted as needed. Online searches suggest that the ratio should be between 0.65 and 0.8. The limits of 0.6 to 0.75 were tested during method exploration. No significant difference was found. Since the Euclidean distance is determined by this ratio, 0.65 is the stricter limit, resulting in fewer matched features. However, due to the cameras' resolution and the blurriness caused by head movements, a less strict ratio, such as 0.75, is recommended. The threshold is, by default, set to five successfully matched features after the ratio test, which is chosen by the lower boundary in the successful send tasks.

7.7.2 Implementation of robot task accomplishment

The implemented task detection model includes the following tasks: "drink," "fill cup," "eat," "scratch," "switch light switch," "brush," "pick object," and "place object." After the presented task selection, the according task has to be performed by the robot. Several research groups are developing solutions for each task. Due to the breadth of related work considering each task, the following discussion about suitable approaches is not exhaustive. We present solutions that demonstrate high-level robot autonomy in completing tasks. If more information is needed to accomplish a task, eye-tracking approaches are presented to continue the modality availability. Some works can perform multiple of the aforementioned tasks. For the sake of conciseness, they are only presented once.

Drinking: The Assistive Gym is a physics-based simulation framework for physical human-robot interaction and robotic assistance, using reinforcement learning to control the robot in various tasks. With this framework, tasks such as drinking can be performed, as demonstrated in Erickson et al. [235]. Due to the large number of integrated robots, this framework is a promising solution for integrating this control. Chi et al. presented a trajectory segmentation approach that can fill a glass of water, hold it near the user, and assist with eating tasks, representing three of the presented tasks [185]. The method by which users can select among these tasks is not specified, underscoring the significance of our framework.

Fill cup: Huang and Mutlu's work presents a user intent prediction-based approach in which users can select individual ingredients via eye tracking to mix a drink. This setup has been used to demonstrate that anticipatory control is faster than reactive control [65]. This approach could help expand possibilities for people with severe disabilities, for example new employment fields.

Eating: Bhattacharjee et al. examined user preferences in a robot-assisted feeding task. They introduced a system that could pick up specific food items via a web interface [15]. Through our implementation of the GazeCursor, interaction with displays and web interfaces is anticipated. Canal et al. presented a setup that can feed people and assist them with dressing themselves [163].

Switch Light Switch: Zhang et al. presented a sophisticated, eye-tracking-based robotic approach capable of picking up and placing objects, toggling switches, watering plants, and pouring water. These tasks were selected based on prompts generated by an LLM model that uses object labels. This approach can be scaled to multiple tasks [22]. Our approach could still be beneficial by enhancing the data density of the given pictograms.

Scratch and Brush: Both Hughes et al. and Dennler et al. present an approach to combing hair [236, 237]. Although this approach is not based on eye-tracking control, the physics are discussed, including different hair types and lengths. No works covering itch-scratching with a robot were found. The literature covers stroking or touching humans as a form of calming social interaction. Scratching is most likely not covered due to the potential for user and participant harm, which is why this task

is usually assigned to caregivers. As Asimov's first law of robotics states, "A robot may not injure a human being or, through inaction, allow a human being to come to harm." [238].

Pick and place: These tasks are becoming established as proof of concept for eye-tracking controls. They lay the groundwork for the aforementioned approaches since most tasks involve reaching for and grasping objects. As presented in Section 7.5.3, the theoretical description was given to obtain coordinates. These coordinates can be used to place an object by calculating the intersection of the surface and gaze vectors. GPD was also mentioned as a solution for generating appropriate grasps [234]. We refer at this point to Marwan et al., who present a review including multiple eye-in-hand approaches for grasping [239].

7.7.3 Limitations and future work

The system must be tested in everyday situations, including scenes with clutter and varying environmental parameters. The test cases were chosen to demonstrate various settings, but were limited to three specific tasks. We documented lessons learned to facilitate use by other researchers. Furthermore, the current integrated, assistive-dataset trained model can detect and track assistive objects with high confidence. However, it is limited to a few classes. The presented logic allows for the use of different models trained on broader datasets to integrate other objects for interaction. Thus, the model can be switched to YOLO models, which are trained on MS COCO and include approximately 80 different classes.

Regarding the use of the framework by people with severe disabilities, an adjustment measure for the robot's field of view should be integrated so that those unable to move their heads can independently adjust the view to align with their gaze. Multimodal inputs have shown promise in opening the framework to people who cannot interact with the provided input modality [64, 240]. Furthermore, interaction and user satisfaction can be improved by refining feedback modalities. In this work, we used auditive cues to indicate successful selection. Other studies have integrated mixed reality, enabling visual cues in the user's field of view [69, 178]. Additionally, LLMs enable more sophisticated auditory feedback, such as information about the object or selected task. These refinements are planned for future work.

7.8 Conclusion

In conclusion, this work introduces a novel eye-tracking-driven shared control framework for assistive robotic arms that effectively addresses challenges related to gaze accuracy in 3D-object detection and complexity related to robustly detecting correct task selection in a system with multiple tasks available. By leveraging task pictograms as fiducial markers and employing a robust feature matching-based communication between the user's eye-tracker scene and a robot-mounted camera, the system enables reliable and intuitive interpretation of task selection without requiring head position tracking. The object's location is found by a kinematic chain through a eye-in-hand configuration. Experimental evaluations demonstrate a high task selection success rate of up to 97.% with a mean response time of 260.3 ms, underscoring the system's robustness and efficiency. While the system reacted adequately in some test cases downfalls were discovered, leading to wrong system reaction. As reaction, counter measures, such as threshold for minimal detected features, were implemented in the code and presented as lessons learned. The open-source framework's modular design and integration of state-of-the-art object detection models facilitate easy adaptation to new tasks and environments. While current tests are limited to laboratory settings and a finite set of objects is available, we presented ways to adapt the system to benefit new approaches. This approach lays a foundation for the integration of real-world applications, that can enhanced independence, usability, and safety for users with severe physical disabilities through autonomous robot assistance.

Chapter 8

Publication VI: Evaluation of Precision, Accuracy and Threshold for the Design of Vibrotactile Feedback in Eye Tracking Applications

Summary: Chapters 1 through 3 highlight the relevance of an alternative feedback method by presenting findings on artificial feedback, the needs of people with disabilities, and tests of integrated feedback methods in Chapters 3, 4, and 7. These works provided users with additional information in the form of auditory and visual cues. In everyday life, we use the visual and auditory channel to perceive our environment. In some social interactions, these feedback cues may not be favorable. For instance, a beeping system in a library could cause discomfort for the user. Receiving additional feedback from the assistive control system over the auditory and visual channels can increase cognitive workload. Consequently, this work investigates the applicability of vibrotactile feedback. This feedback method uses touch. Discrete vibration cues inform users about certain events, such as successfully selecting a task.

The first study determined perception thresholds at the temple. The second study measured accuracy and precision errors with and without the vibrational stimulus to estimate its impact on the device. No significant differences were found. These results led to the conclusion that the vibrotactile system can be used as a feedback method. Restrictions and implementation suggestions were presented to facilitate this approach in a system designed for people with disabilities.

Contribution: Anke Fischer-Janzen (former Anke Fischer) was the primary author of this publication. She was responsible for the conception, implementation, evaluation, and preparation of the manuscript. Lukas Stiglemeier, Philipp Gawron, Thomas M. Wendt, and Kristof Van Laerhoven contributed to this work through supervision, methodological discussions, and manuscript review.

Evaluation of Precision, Accuracy and Threshold for the Design of Vibrotactile Feedback in Eye Tracking Applications

Anke Fischer
Offenburg University of Applied Sciences

Thomas M. Wendt
Offenburg University of Applied Sciences

Lukas Stiglmeier
Offenburg University of Applied Sciences

Philipp Gawron
Offenburg University of Applied Sciences

Kristof Van Laerhoven
University of Siegen

Abstract

Novel approaches for the design of assistive technology controls propose the usage of eye tracking devices such as for smart wheelchairs and robotic arms. The advantages of artificial feedback, especially vibrotactile feedback, as opposed to their use in prostheses, have not been sufficiently explored. Vibrotactile feedback reduces the cognitive load on the visual and auditory channel. It provides tactile sensation, resulting in better use of assistive technologies. In this study the impact of vibration on the precision and accuracy of a head worn eye tracking device is investigated. The presented system is suitable for further research in the field of artificial feedback. The vibration was perceivable for all participants yet it does not produce any significant deviations in precision and accuracy.

8.1 Introduction

In the past few years, the research in the field of eye tracking based controllers for robot interaction increased, as shown by different authors and their work [241, 19, 51, 47, 52, 104]. Eye tracking yields advantages against other non-verbal or touch-based systems such as an easy setup and a good accuracy. Challenges of near-infrared eye trackers are the limited range of motion, especially when using stationary eye tracking devices [104]. Wearable, wireless eye tracking systems represent a solution.

In this context, the research of user experience or improvement measures such as artificial feedback is scarce. Vibrotactile feedback is a commonly applied method to simplify the usage of electric protheses or smart devices [242, 243]. For this, the feedback stimulus has to be well adjusted, perceivable and must be easy to assign to the corresponding event [137]. Current studies have shown that placement is crucial for a well-matched feedback [244, 245, 246, 134]. Applications for stimulus locations such as the head and chest were tested in guidance vests, belts, helmets and glasses [245, 246, 247, 134]. The results showed that tactile feedback is improving the completion time of gaze gestures depending on the duration of the stimulus [248]. Further, the cognitive load of the auditive and visual channel was reduced [246].

Implementing the feedback method in eye-tracking glasses could reduce set-up time for caregivers and users by replacing a separately mounted feedback system. In this work, possible influences on the sensor technology of the eye tracking glasses caused by the vibration are investigated. In the first study, perceivable vibrotactile thresholds in the temple area were determined to find lower boundaries for the feedback. The second study investigated variations in the precision and accuracy between deactivated and activated vibrotactile feedback of a head-worn eye tracker. A first insight of the user experience of this setup is given.

8.2 Methods

Two studies with ten persons were performed. The results were utilised to determine an upper and lower vibration perception threshold. The first study was designed as an empirical threshold measurement in which a piezo element was used as vibration stimulus in the temporal region. In the second study a vibration coin motor created the stimulus on a pair of eye tracking glasses. Precision and accuracy were evaluated based on the recorded data.

The participants (6 males and 4 females) were in average 34 years old (SD 10 years) and agreed to participate in both studies. None of the participants stated medical conditions concerning eye muscles or tactile perception. No medical conditions such as headaches, nausea or dizziness were stated by the participant in the duration of the study. The study was reviewed and approved by the Ethics Counsil of the University of Siegen. All participants read, agreed and signed the informed consent forms before conducting the studies.

8.2.1 Vibration perception threshold of the temporal region

In the first study, a piezo element (PK2FVP2, ThorLabs) was characterised with a laser vibrometer (OptoMET Dual Sense I) and used as stimulus. Two frequencies were tested with this setup. 54 Hz were applied for the exploration of user experiences with vibrational feedback at the head by Myles and Kalb [246]. The second applied frequency of 140 Hz corresponded to the created vibration of the coin motor used in the second study. The linear amplitude change for the threshold measurement was between 2.4 µm and 5.2 µm (SD $\pm$0.05 µm). Errors due to bias were minimised by not interrupting participants when they exceeded the stated threshold range.

The piezo element was placed so that the expansion force acted vertically on the skin. A customised 3D-printed housing was designed to adjust the position on the head. An adjustment screw allowed the position of the piezo element to be regulated to ensure contact with the skin. The participants were told to state if the stimulus was perceivable with "yes" or "no" answers. The amplitude was adjusted by setting the voltage in 0.1 V steps corresponding to a difference in amplitude of 60 nm. 50 measurements were conducted for each participant and frequency. The on-time of the stimulus lasted one second to reduce the risk of adaptation. The transformed "up/down" Method introduced by Wetherill and Levitt with "3 up/1 down" was used to determine the threshold ([249], p.124-125). The results were evaluated with a psychometric function which was fitted to the data by using a generalised linear regression model with MATLAB, as presented by Wichmann and Hill [250]. The standard deviation was used to evaluate the results regarding the reliability of the participants statements.

8.2.2 Measurement of accuracy and precision in an eye tracking task with vibrotactile stimulation

The apparatus consisted of an eye tracking device (Tobii Pro Glasses 2, Tobii Pro) and a 3D-printed casing enclosing the vibration coin motor (Type 310-122, PMD). To ensure an error-free operation of eye tracking systems, the environment must meet certain criteria. Eye tracking devices are sensitive to changing illuminance of the scene, calibration errors and, in the case of wearable systems, fit to the head. These affect the accuracy and precision of eye movement detection [251, 252]. Therefore, illuminance was set to 480 lux through adjustable lights to minimise illumination errors. In total, 13 visual markers 65 and one reference marker were presented centrally in front of the participant's eye level, as shown in Figure 8.1. The spacial distribution was adopted as stated by Thibeault et al. [251]. The markers were arranged in circles with angles of 7, 15 and 23°to the center marker. Each marker had a diameter of 20 mm (adapted from [251]). The reference marker had a diameter of 30 mm.

The participant was sat in front of the marks with a distance of 1 m. Shorter distances resulted in uncomfortable head postures, masking of the markers by the glasses and reduced detectability of the outer markers by exceeding the scene camera image. If necessary, the glasses were adjusted with different nose pads and calibrated as stated in the manual [253]. Participants were allowed to move their head between measurements. The head position and distance were adjusted with a chin rest, to minimise a variation of the head position and the resulting errors. This variation is shown in Figure 8.1 by two exemplary selected eyeglass planes that are shifted in relation to each other, each representing one measurement. In this case, the tilt to the left would lead to a non-ecognition of the outer right point (blue-coloured camera view). Since the gaze angle is measured in pixels in relation to the scene camera view, a slip or tilt of the glasses can result in varying gaze angles for the same marker.

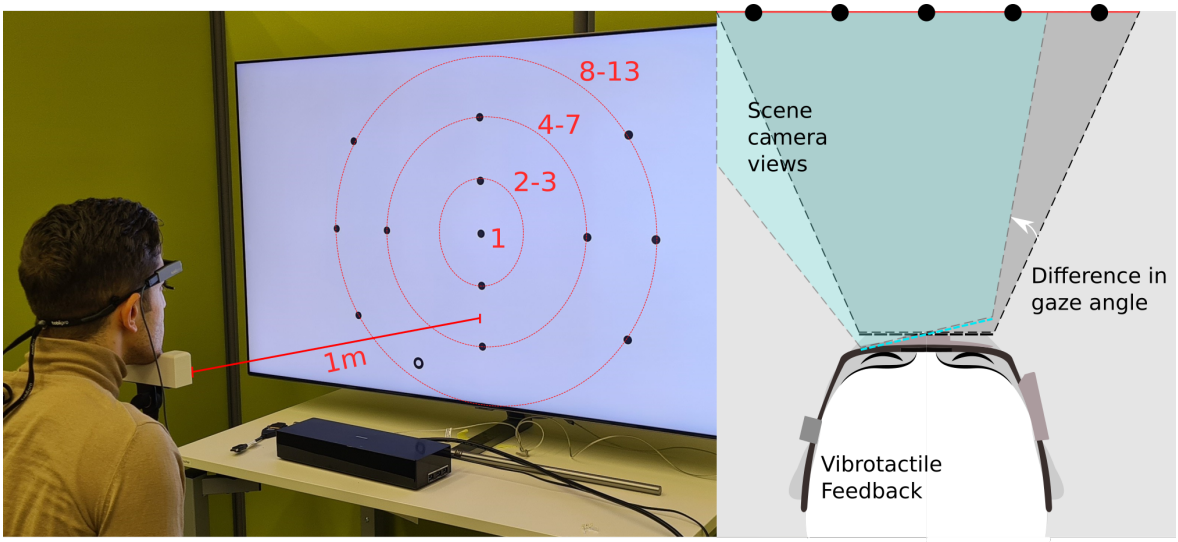


Figure 8.1: Left: Study setup with participant. All red markups weren't visible for the participant. The distribution of the markers was adapted from Thibeault et al. [251]. Right: The scene camera view, shown as highlighted area between participant and markers, represents the room of all available gaze angles. Diagonal lines in black and blue show examples for a tilted level of the glasses.

The vibrotactile stimulus was set to a frequency of 140 Hz with a resulting amplitude of 75 µm. All participants stated that the vibration was perceivable in the temporal region. Eight measurements were conducted, alternating between deactivated (Condition A) and activated (Condition B) vibrational stimulus to reduce the impact of habituation effects. A preliminary study showed that more than eight measurements lead to an increased fatigue stated by the participants. In all measurements, each marker was focused for three seconds to generate data sets of at least 300 data points. After the measurements, a short survey was conducted regarding the sensation of the glasses and the casing as well as the vibrotactile feedback. The vibration was audible within these settings. Therefore, the perception of the auditory feedback was asked to distinguish between the sensation of the vibrational and auditory feedback. Since learning behavior or evaluation of the cognitive load wasn't focused in this study, no further measures were taken to prevent a perception of the auditive feedback. The participants had to indicate the sensation on a scale from 1 (very pleasant) to 5 (very unpleasant). If the sensation was stated by a level greater 3, further questions to the occurrence of this decision were asked.

The raw data was exported with Tobii Pro Lab for further analysis with MATLAB. The program flow is shown in Figure 8.2. The arithmetic mean was calculated for each marker and each participant. The precision was calculated as stated by Thibeault et al. [251] as RMS of the standard deviation in x and y direction. Accuracy was calculated by determining the length of the vector from the position of the

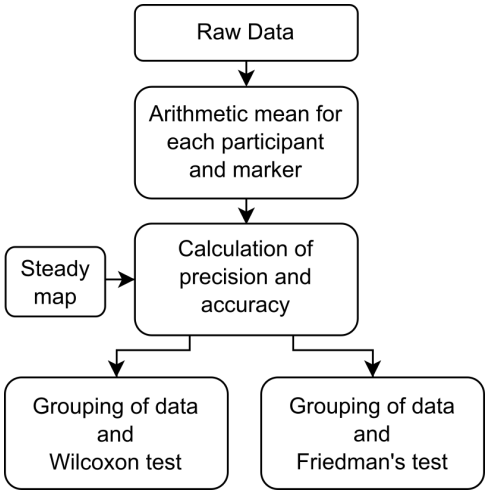


Figure 8.2: Program flow of the data analysis. Precision and accuracy were determined and compared with a steady map. Significance between conditions and measurements were calculated with the Wilcoxon rank sum test and Friedman's test.

arithmetic mean marker to the corresponding marker in a steady map. Marker M1 of the steady map was estimated to represent the center of the eye tracking glasses scene camera view. The precision and accuracy were investigated on significant differences between the AB-conditions. The data of the AB-study design was not normally distributed. Therefore, a Wilcoxon rank sum test was executed with the grouped data of all participants into the corresponding condition. A Friedman's test with following Dunn-Bonferroni-Test was applied to determine differences between measurements. The vibration used in this test can cause the glasses to slip, which is expected to affect accuracy and precision. Therefore, a sliding window and a scaling factor were calculated as a correction factor and applied on the data for the accuracy to estimate systematic errors. Due to possible movement of the participants between the measurements, the significance tests were repeated after using the sliding window and scaling factor.

8.3 Results

With regard to comparable literature on threshold measurements at the temple, the results showed that vibration perception was dependent on the age of the participants and the applied frequency. Stuart et al. stated amplitudes of 28 µm to 107 µm depending on the participants age, whereas Myles and Kalb stated thresholds around 2 µm at varying frequencies [246, 254]. The vibration perception threshold measurement in the first study was carried out with all participants. In five cases, the threshold couldn't be determined, due to a non-sufficient maximum amplitude of the system. The participants who perceived the vibration with an amplitude lower than the maximum of 5.7 µm were in average 29 years old. Participants who couldn't perceive the stimulus were in average 43 years old. The standard deviation shows in comparison to the measured amplitude high deviations between 0.13 µm and 0.39 µm. This can be attributed to a small step size and thus inconsistent answers from the participants.

In the second study, the data sets for each marker containing between 61 and 590 values were evaluated. As stated above each point was fixated for 3 seconds to generate at least 300 data points depending on the sampling rate and fixation duration. Less than 300 data points were recorded since no data is captured, if the gaze direction exceeds the scene camera view. The corresponding markers were located exclusively in the periphery. These markers were excluded from the evaluation as soon as fewer than 50 values were recorded, leading to 26 excluded data sets of a total of 1040.

Table 8.1: M1 describes the centred marker and M8–M13 the outer markers (see Fig. 1). All data are given in millimetres except the scaling factor, which has no units and for which average improvement is shown as a percentage. (a) Accuracy and precision calculated with the steady map and without correction factors. The means of precision and accuracy are stated for the conditions with (w. FB) and without vibrotactile feedback (w/o FB). (b) Effects of the sliding window and scaling are shown by an average improvement rate for each participant, improving the error between 32 % and 81 %.

P.	Prec. M1		Prec. M8 - M13		Acc. M1		Acc. M8 - M13	
	w/o FB	w. FB	w/o FB	w. FB	w/o FB	w. FB	w/o FB	w. FB
3	9.86	6.82	16.80	21.03	75.58	105.17	93.01	154.27
4	13.54	6.77	24.12	17.07	155.46	152.97	165.96	185.95
P.	Average	Slide in	Slide in	Scaling	Acc. Slide M1		Acc. Slide M8 - M13	
	Improvem.	x-direct.	y-direct.		w/o FB	w. FB	w/o FB	w. FB
3	32.65	30.76	-76.69	0.83	9.96	9.21	70.70	198.39
4	81.89	-108.71	-99.77	1.12	16.60	4.92	47.35	47.02

Table 8.1 shows the resulting precision and accuracy for different groups of makers at the example of participant 3 and 4. These participants were selected because their data showed a minimum and maximum rate of improvement by the applied correction factor. M1 describes values taken when fixating the centered marker. The values given for M8 - M13 describe the arithmetic average of the parameters for the outermost markers with a radius of 23°.
The scaling and sliding factors are given, with the resulting accuracy values in the lower part of the table. A minimum of 4.52 mm in accuracy is reached. Between conditions no significant difference was found. The errors in accuracy between the inner (M1) and the outer markers (M8 - M13) increase the closer the marker is located to the periphery of the camera view. To optimise the position of the markers, they have to be shifted by 189.59 mm and scaled by a factor between 0.71 and 1.12 with the sliding window method. This correction reduces the calculated error by up to 81.89 %.

In Figure 8.3 the weighted average of the accuracy for each participant and each condition is shown after applying the sliding window and scaling factor. For some participants, lower accuracy for the condition with feedback can be seen, such as in the data of participant 3 and 9. Although this implies that vibration has an impact on accuracy, the data from participant 10 indicates the opposite. The statistical analysis showed similar results for all participants. No significant differences between the grouped precision and accuracy data was found by using the Wilcoxon test. There are differences between measurements (Friedman test $p<0.05$). A conducted post-hoc test (Dunn-Bonferroni-Test) showed no significance since the alpha-value was reduced.

The participants were asked to evaluate the comfort of wearing the glasses and the mounted casing of the vibration motor, the applied vibration and the auditive cue resulting from the stimulus. The perception of the eye tracking device was rated as pleasant in all measurements (mean 2.07, SD 0.53). The perception of the vibrotactile feedback was rated as neutral to unpleasant, with a mean of 3.34 (SD 0.94). The auditory perception was rated with 2.96 (SD 1.07). The participants tended to stay at the same statement and adjust it with a maximum of one point in between all measurements, leading to the assumption that the duration of the test was adequate. Reasons for statements worse than 3 included that the vibrotactile feedback was too strong and the vibration is noticeable over the whole frame. Main

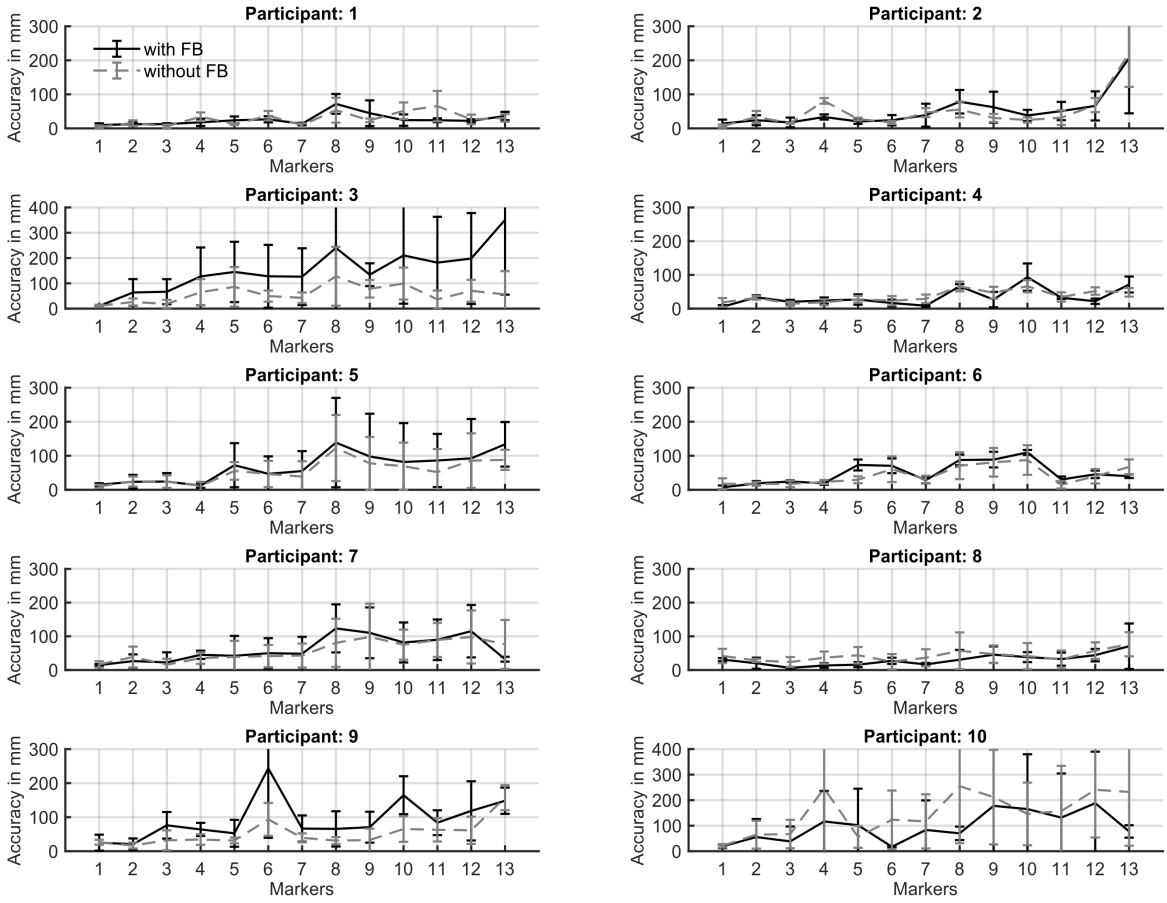


Figure 8.3: Weighted average for both conditions (activated vibrotactile feedback (black), deactivated vibrotactile feedback (green)). The distribution of the markers are shown in Figure 8.1.

reasons for an unpleasant auditive feedback experience were the volume and pitch of the motor.

8.4 Discussion

The threshold values determined in the first study can be compared with the results of Myles and Kalb for the temporal region, as threshold values around 2 µm were found [246]. Since thresholds vary regarding to age, type of stimulation, and skin and medical conditions, the study setup and design is shown to be applicable for further measurements. Higher thresholds for the frequency of 140 Hz were shown for all participants. Age is one reason why the threshold tends to rise, as Stuart et al. found [254]. With regard to the group exceeding the upper limit of the amplitude values, these results can be agreed upon. On average, an age difference of 14 years lies between both groups. Regarding the setup and procedure, the step size of the threshold measurement should be increased to lower the standard deviation. The vibration of the piezo element was audible when activated. Due to the bone conduction of the sound waves, it was not possible to suppress the perception of the noise. Since similar results were acquired as in current literature, the results are still taken as reliable. In consideration of vibrotactile feedback design in head worn eye tracking applications, a lower range couldn't be determined. As an indication, the amplitude has to exceed the used 5.7 µm by a factor of at least three in further studies. This is estimated by the conversion of the presented data by Stuart et al. [254].

In the second study, the statistical analysis has shown that no significant differences could be found between the AB-conditions. The occurring acceleration of the glasses caused by the vibration was measurable by the built-in accelerometer. An effect on the precision and accuracy could not be found. Since the markers had a diameter of 20 mm, the errors in precision only surpass an error greater 20 mm in the periphery. This can be shown for all participants for M8 - M13 and in one case at the centered marker. In Table 8.1, examples are shown by the data of participants 3 and 4. These deviations can be caused by outliers, since the data was not filtered to receive more reliable data. Higher errors can be reasoned by an insufficient detection of the pupil in the periphery, as shown by Thibeault et al. [251].

Accuracy errors found in this study rise up to a maximum of 290 mm in comparison to a steady map. Such errors can occur by different effects. The distance to the board may vary slightly, as well as the position of the glasses with each measurement, since the participants were allowed to move the head between measurements. It is assumed that such deviations occur from systematic errors. For example, they are caused by calibration errors. These repeat for all measurements. Sliding window and scaling can minimise such errors. Other sources of error can influence the result. If the resulting tilt angle of the head and distance to the board is calculated from the maximum sliding window and scaling factor, a tilt of 10°around the z-axis and a variation of distance with +12 cm and -29 cm derives. Since the head position may differ slightly from the estimated position of the static map, it is conceivable that the sliding window factor is composed of the head position and calibration error. The expected distance variation caused by a wrong head position is estimated to be less than one centimeter. It is assumed that another reason is the tilt between the level of the glasses and the level of the markers. Because of that, some markers appear closer and bigger in the scene camera. In the statistical analysis of the accuracy, significant differences between the conditions were found, when comparing all measurements with each other by using the Friedman's test. The Dunn-Bonferroni-Test used as post-hoc-test showed no significance for each measurement. As a reason for these outcomes, it was analyzed that a very low statistical power exists ($\beta < 0.2$), due to a small difference between both conditions and the sample size of 10 participants. As the survey on the satisfaction and comfort of the feedback revealed a slight dissatisfaction with the frequency and amplitude mentioned, no measurements were carried out with further participants in order to improve the statistical power. One measure for increased comfort is to reduce the strength of the vibration. This will also reduce the impact on the glasses.

Eye-tracking-based controls can be used with the presented setup if the object is near the centre of the field of view. Under this condition, the error in precision and accuracy is minimised. The results show that the measured values scatter in accuracy within a range of 2 cm at a target distance (between participant and screen) of 1 m and that in most cases accuracy improved with the sliding window and scaling factor. Regarding the errors in accuracy, error handling is needed for the continuous use of eye tracking controls. Different approaches suggest the use of gaze gestures or longer dwell times to improve reliability [248]. These methods are limited when we consider challenges such as slipping of the glasses. Both eye tracking methods can be combined and supported with the use of artificial feedback. Assistive technology users could benefit from artificial feedback.

Considering the design of the feedback, the amplitude should be lower than 75 µm, since the feedback was rated as too strong in the questionnaire. Similar occurrences were also found in the results of Myles and Kalb [246]. Further, Myles and Binsel [247] stated frequencies above 150 Hz are not recommended, due to causing headaches, nausea and other medical conditions. It could not be precisely determined whether these complaints were triggered by the vibrotactile feedback, auditory feedback or both. For this reason, the frequency should be adjustable by the participant and lower 60 Hz to reduce auditory stimulus and increased user satisfaction. This measure was also confirmed by Stuart et al. [254]. Finally, regarding the design of the feedback, the placement of the stimulus should be

reconsidered. For example, the system should be placed behind the ear to reduce the vibration of the glasses frame.

8.5 Conclusion

The application of vibrotactile feedback was evaluated for a head worn eye tracking device. Threshold measurements and empirical surveys in two studies indicated feasible intervals for amplitudes of vibrotactile feedback at the temple. No significant difference on the precision and accuracy of measurement data with deactivated and activated vibrotactile feedback could be found. As shown in the sliding window and scaling factor, the error can be reduced to a few millimeter. Regarding people with reduced motor control, methods such as gaze gestures should be considered or calibration techniques for a continuous minimisation of the error while using the glasses. The applied vibration was characterised with an amplitude of 75 µm and a frequency of 140 Hz. The survey showed that the high amplitude was rather uncomfortable for the participants. Based on these results, it is assumed that vibration has no effect on gaze detection. However, weaker vibration should be used for user experience and learning behaviour studies. It is part of future works to evaluate, if artificial feedback supports the control of smart wheelchairs and robotic arms for handicapped people.

Chapter 9

Conclusions

9.1 Summary and synthesis of findings

This section discusses the scientific contributions presented in Section 1.3 to answer the research questions posed in Section 1.2. The contributions from each publication are synthesized to provide a broader understanding of the design of eye-tracking-driven robotic arm controls.

Research Question 1.
What contributes to the user experience of eye-tracking-driven controls for assistive robotic arms?

The acceptance of these controls can be influenced by different characteristics. Contributions 1.a to 1.g resulted from the studies conducted to evaluate these influences in Chapters 3 and 4. In this context, usability, functionality, and accessibility were investigated. This topic was explored through a review of existing literature and an investigation of the routines and requirements of people with disabilities living in shared homes.

It was found that the system must not limit the use of other devices, like ventilators or catheters. Large, heavy, or stationary devices limit the system's usability due to increased discomfort from wearing heavy equipment for extended periods of time and limited mobility. Large devices obstruct daily routines, such as toileting, by restricting mobility in collaborations between users and caregivers. Higher setup and calibration times may result in rejection because caregivers often lack time for these tasks alongside their daily work. On the other hand, these systems are more accurate, which can improve functionality.

The input modality and underlying control strategy must fit the user's abilities. Users will reject systems that they cannot properly control. An analysis of works comparing different input modalities revealed distinct benefits. The results showed that, when correctly implemented, eye-tracking is showing benefits in contrast to other input modalities. The length of the fixation needed to trigger actions in dwell-time approaches significantly influences minimizing the "Midas touch" problem. Additionally, each study identified limitations related to the integration approach when using eye-tracking, highlighting the need for in-depth planning, testing, and integration of features that fit users' needs.

The importance of including potential users and integrating artificial feedback to communicate the system's response to the user was discussed. Two lists were presented that investigated the characteristics of existing systems and the objective and subjective measurement parameters. These lists aim to facilitate future developments by providing a concise overview of meaningful result

representation. Lastly, ethical considerations in the deployment of robotics in assistive living were discussed. Since the presented work and adjacent literature mostly rely on camera usage, addressing ethical concerns is essential to ensure user privacy.

In conclusion, the technical approach itself determines user acceptance, as it may limit functionality. The characteristics and needs of the user group, a suitable input modality, and the environment affect the system's accessibility. The tasks that the robot can accomplish, the artificial feedback modalities, and the anticipated human-robot interaction determine its usability.

Research Question 2. What usability and functionality challenges arise with object detection-based eye-tracking control?

The aforementioned influences on usability and functionality were considered when developing the initial prototype control used in the Wizard of Oz studies. Contributions 2.a through 2.c addressed the research question. Object detection was chosen as the approach so that the resulting bounding boxes could be used as interactive buttons. The robot selects the task associated with the object and carries it out. This action is triggered when the user gazes continuously at the object.

The research conducted unveiled limitations in correctly detecting objects, which led to challenges in robustly selecting them. Object geometry generates bounding box overlap. The functionality as buttons was greatly limited. The data showed that user intent influenced gaze behavior, leading to the object's fixation near the borders of the bounding boxes. This behavior, together with low accuracy, resulted in gaze movements that were tracked outside the bounding box, which prevented the selection of the object. The gaze direction, as detected in relation to the periphery of the eye-tracking glasses and the object size, limited the system's functionality due to the small button sizes and the reduced accuracy.

Integrating new tasks, which would make the system more scalable and beneficial in everyday life, is limited in cluttered scenes due to objects being close to each other and the aforementioned accuracy issues. The risk of selecting the wrong object, and hence the wrong task, increases. Additionally, integrating the robust selection of a higher number of tasks is only partially possible due to ambiguous object interpretation relevant to the task. The necessity of integrating multiple tasks was demonstrated through interviews with people with disabilities, their families, and their caregivers, as well as an analysis of preferred daily tasks. Without a framework that distinguishes between actions such as drinking from a glass and filling the same glass with water, integrating multiple tasks would result in loss of agency and eventually lead to rejection of the system. Therefore, a new strategy that answers RQ 3 will be presented in the next paragraph.

Research Question 3. How can robustness and scalability in task selection be addressed?

Contributions 3.a through 3.f support answering this research question. The approach resulted in a pictogram-based method that improved user understanding and proposed using them as fiducial markers to increase information density. The trained model was integrated into a framework that represents the eye-tracking-based control. This framework solved robustness issues in pictogram detection by incorporating a fallback mechanism that disregards incorrect predictions. Many authors have noted that limited accuracy is an issue when estimating the location of the desired object in eye-tracking controls. I overcame this challenge by transmitting object information from the eye-tracking data to the robot camera using a feature-matching approach. Thus, the head position does not need to be known. This approach enables free head movement and reduces physical stress for the user.

The functionality was evaluated through a study investigating different use cases users could encounter. For instance, the robot can be mounted in various locations, such as on an electric wheelchair or a table for feeding tasks. Different illumination settings and backgrounds were evaluated. A suitable feature-matching algorithm was chosen to enhance functionality. Due to the lightweight setup of only two cameras and the absence of additional sensors (e.g., IMUs or movement-tracking markers), the anticipated accessibility requirements were met.

Although usability must still be investigated, initial measures have been taken to ensure the anticipated benefits for users, such as performing various daily tasks and the ability to continuously select tasks that promote independent living. This approach outlines one suitable possibility to achieve robust and scalable task selection. Other approaches can lead to similar results. A comparison with other approaches was outside the scope of the presented studies.

Research Question 4. What conditions are suitable for applying vibrotactile feedback to the eye-tracking glasses?

This evaluation was deemed necessary since the auditory and visual channels already face a heavy workload in everyday life. Since gaze is used for control and listening is used for conversation, additional cues could overwhelm the user. Auditory cues could lead to rejection of the system because they could be disruptive in quiet settings and disrupt social norms, similar to the occurrence of ringing smartphones in libraries. In conclusion, this paragraph elaborates on the research question with contributions 4.a to 4.c. The conditions in Chapter 8 were evaluated with a vibrotactile feedback prototype. Although no significant difference was found between the baseline measurement and the measurement with activated feedback, concerns were raised because the size of the visual marker exceeded accuracy errors. The application of vibrotactile feedback may be restricted in the eye-tracking-driven system due to the minimum pictogram size of 2 cm. This size was chosen because fixation on the pictogram's center will still be tracked inside the bounding box, even according to the manufacturer's specifications on accuracy error. Therefore, an increased error would cause the selection process to track gazes outside the bounding box. During the evaluation process, different application locations were suggested, such as behind the ear, based on participant dissatisfaction and reduced vibration transfer to the front of the glasses. Under these conditions, vibrotactile feedback can be used as an artificial feedback method in eye-tracking-driven controls. Further investigations must be conducted when using a different type of eye-tracking glasses.

The research focus on eye-tracking-driven controls for robotic arms has shifted from simply creating such controls to profoundly investigating and improving human-robot interaction in these controls. User experience is a high-level concept influenced by usability, functionality, and accessibility. This doctoral thesis presented works investigating different characteristics that influence the user experience. The work revealed influences that would severely limit functionality in practice of object-oriented selection methods and showed an alternative way of addressing the core challenge of robust selection. The synthesis of the found insights from usability evaluation, object detection, eye tracking, and artificial feedback establishes design principles for robust and user-oriented eye-tracking-driven controls. This thesis provides a structured foundation for the next generation of eye-tracking-driven controls for robotic arms that must balance technical robustness with human-centered design.

9.2 Limitations

This work primarily focused on eye-tracking-driven controls for assistive robotic arms. Without the limitation on this input modality, a vast amount of related publications regarding user acceptance in robotics, various robotic types, input modalities, shared control approaches, and AI developments can be found. Thus, the scope of this work was narrowed to present valid, generalizable data on robotic arm usage with eye-tracking-driven controls.

Merging the investigated control components, such as artificial feedback, the eye-tracking framework, and robot control, is part of future work. No conclusion can be made regarding the resulting user acceptance of the complete framework, although the individual modules were found to be applicable and indicate a favorable user experience. The following section presents challenges that emerged from the synthesis discussed in Section 9.1 and that may result in a adverse user experience.

The system was designed to enable users to move in the task selection process and reduce the effects of limited accuracy caused by slippage of the glasses. As indicated in Chapter 7, the task is not sent when the confidence level is too low for correct transmission. This instance occurs when only a few features are tracked. In order to avoid repeatedly measuring a low-confidence case, the perspective must be changed by either moving the head or the robot camera in order to reduce occlusions of objects. Regarding the system's accessibility, this solution may not be feasible for people with tetraplegia due to their limited ability to move their heads. Integrating the movement into the robot to slightly vary the position of the robot-mounted camera might be feasible. Since such an endeavor would require extensive evaluation and integration of new features, it was considered outside the scope of this work. This represents an important limitation for real-world use and leaves open questions about the user experience of people with disabilities.

Object detection models are not naturally robust because they are trained on specific data. Each representation of the real world in a dataset is only an estimated representation due to factors such as illumination, backgrounds, camera angles, how objects are represented, and the aging of objects and their surroundings. Therefore, measures such as the fallback mechanism were implemented to mitigate the impact of these factors on the proper control functioning. In the context of Chapter 6, robustness was defined as the ability to perform multiple tasks, as needed in real-world settings.

9.3 Future Work

This section will highlight research topics related to eye-tracking-driven controls that can benefit from or contribute to the development and investigation of other technologies.

The presented publications resulted in an eye-tracking-driven control for robotic arms, building on previous research in usability and accessibility. As discussed in Chapter 7, questions remain regarding these two topics. Investigating the influence of different artificial feedback methods can demonstrate their benefit to this system. This work investigated vibrotactile, auditory, and visual feedback within a specific scenario. Future work can focus on whether these findings are generalizable to different scenarios, which are currently investigated in the context to mixed reality approaches. People with disabilities should be included in the future to assess the system and evaluate the previously raised points on usability and accessibility. While it is reasonable to hypothesize that the system provides added value for this group, the results must be investigated objectively.

In the future, eye-tracking-driven controls will greatly benefit from new developments in the fields of Vision Language Models (VLM) and Large Language Models (LLM). Text-based analysis provides a new opportunity to represent tasks in images and present the system additional information. Another way AI can be used in these controls is to adapt the level of autonomy to users' preferences. These developments could lead to the adaptation of gaze behavior analysis, raising open questions about how to integrate such an adaptation to optimize the user experience.

Integrating additional tasks with the robot is an ongoing process. As part of the HIRAC project, the "pick and place," "drink," and "pour water" tasks were implemented. They need to be connected to the control system and undergo functionality testing. Research on daily tasks revealed that certain tasks, such as scratching, have never been performed by a robot, possibly due to ethical concerns. These borderline cases offer new potential applications and perspectives on the use of robots in ethically complex situations and assistive living.

List of Figures

3.1 Comparison bright and dark pupil detection . . . 14
3.2 PRISMA2020 identification and screening process . . . 19
3.3 Characterstics overview of included works . . . 25
3.4 Comparative Study overview of input modailities . . . 29
3.5 Search term analysis. . . . 34

4.1 Application and requirement analysis process. . . . 38
4.2 Example of the shared apartment . . . 39
4.3 Use of robots in different environments . . . 44

5.1 Wizard of Oz design . . . 55
5.2 Interactable objects and tasks . . . 56
5.3 Question distribution in the online study . . . 58
5.4 Response distribution of the online study . . . 61
5.5 Task classification in the online survey . . . 62
5.6 Attention visualization during task selction with heat map . . . 63
5.7 Correlation between object size, missed hits and dwell time . . . 64
5.8 View point heat map evaluation . . . 65

6.1 Control process . . . 78
6.2 Upper row: Initial pictogram design, Lower row: Redesigned pictograms. . . . 81
6.3 Normalized confusion matrix for test image size of 1280 px. . . . 84
6.4 Model performance with varying distance between pictograms and camera. . . . 85
6.5 Use cases, Overlap of bounding boxes. . . . 86

7.1 Computation process of the framework . . . 97
7.2 Feature matching test case description . . . 100
7.3 Feature matching results for selected algorithms . . . 102
7.4 Program flow for robust information transmission . . . 105
7.5 Framework description . . . 107
7.6 Test objects and attached pictograms . . . 108

8.1 Study setup and scene camera view . . . 116
8.2 Program flow of the data analysis . . . 117
8.3 Result overview: Accuracy as weighted average . . . 119

List of Tables

3.1 Research Questions and motivation 13
3.2 Databases used in the identification phase. 16
3.3 Inclusion and exclusion criteria 17
3.4 List of articles included 17
3.5 Standardized questionnaires for subjective evaluation of assistive robotics. 32

4.1 Feedback key points regarding the usage of Pepper. 40
4.2 Functional and non-functional requirements for robot application in assisted work. . . 42

5.1 Wizard of Oz presentation 54
5.2 Breakdown of topics asked in the survey and related questions. 57
5.3 Distribution information of task selection 63
5.4 Overview of tasks mentioned by participants. 67
5.5 Questions asked in the online survey 71
5.6 Participant statements in the hands-on study 73

6.1 Overview of task interpretation 82
6.2 Model performance after training 83
6.3 Advantages and disadvantages of key contributions. 87
6.4 Overview of the answers available in each question and response distribution. 91

7.1 A comparison of publications on FDDs and FM algorithms based on OpenCV. 99
7.2 Algorithm comparison results 101
7.3 Summary of the evaluation test cases 110

8.1 Precision and accuracy overview with offset and scaling factors. 118

Acronyms

ADL Activities of Daily Living.

AI Artificial Intelligence.

ALS Amyotrophic Lateral Sclerosis.

ARA Assistive Robotic Arm.

BCI Brain Computer Interface.

EEG Electroencephalography.

EOG Electrooculography.

hBCI hybrid Brain Computer Interface.

HMI Human Machine Interaction.

HRI Human Robot Interaction.

iEEG intercranial electroencephalography.

IOG Infrared Oculography.

MARG Magnetic, Angular Rate, and Gravity.

QoL Quality of Life.

SSVEP Steady-State visually Evoked Potential.

TCP Tool Center Point.

VOG Video Oculography.

References

[1] Destatis. “Press release of 22 June 2022: 7.8 million severely disabled people living in Germany”. 2022. URL: https://www.destatis.de/EN/Themes/Society-Environment/Health/Disabled-People/_node.html.

[2] 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”. In *Technologies* 9, 8 (2021), 1–26. DOI: https://doi.org/10.3390/technologies9010008.

[3] A. Calvo, A. Chiò, E. Castellina, F. Corno, L. Farinetti, P. Ghiglione, V. Pasian, and A. Vignola. “Eye Tracking Impact on Quality-of-Life of ALS Patients”. In *Computers Helping People with Special Needs*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2008, 70–77.

[4] 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, 6999–7006. DOI: https://doi.org/10.1109/ICRA46639.2022.9811960.

[5] A. Hagengruber, G. Quere, M. Iskandar, S. Bustamante, J. Feng, D. Leidner, A. Albu-Schäffer, F. Stulp, and J. Vogel. “An assistive robot that enables people with amyotrophia to perform sequences of everyday activities”. In *Scientific Reports* 15, 8426 (2025). DOI: https://doi.org/10.1038/s41598-025-89405-2.

[6] A. Nanavati, V. Ranganeni, and M. Cakmak. “Physically Assistive Robots: A Systematic Review of Mobile and Manipulator Robots That Physically Assist People with Disabilities”. In *Annual Review of Control, Robotics, and Autonomous Systems* 7 (2024), 123–147. DOI: https://doi.org/10.1146/annurev-control-062823-024352.

[7] H. O. Edughele, Y. Zhang, F. Muhammad-Sukki, Q.-T. Vien, H. Morris-Cafiero, and M. Opoku Agyeman. “Eye-Tracking Assistive Technologies for Individuals With Amyotrophic Lateral Sclerosis”. In *IEEE Access* 10 (2022), 41952–41972. DOI: https://doi.org/10.1109/ACCESS.2022.3164075.

[8] L. Y. C. Loke, D. R. Barsoum, T. D. Murphey, and B. D. Argall. “Characterizing eye gaze and mental workload for assistive device control”. In *Wearable Technologies* 6 (2025), 1–22. DOI: https://doi.org/10.1017/wtc.2024.27.

[9] D. Zhu, T. Gedeon, and K. Taylor. “Head or Gaze? Controlling Remote Camera for Hands-Busy Tasks in Teleoperation: A Comparison”. In *Proceedings of the 22nd Conference of the Computer-Human Interaction Special Interest Group of Australia on Computer-Human Interaction*. ACM Other conferences. New York, NY: ACM, 2010, 300–303.

[10] Y. Cao, S. Huang, S. L. Sørensen, Y. Yamakawa, and M. Ishikawa. "A Wearable Real-Time 2D/3D Eye-Gaze Interface to Realize Robot Assistance for Quadriplegics". In *IEEE Access* 13 (2025), 155259–155276. DOI: https://doi.org/10.1109/ACCESS.2025.3603617.

[11] A. Belardinelli. "Gaze-based intention estimation: principles, methodologies, and applications in HRI". In *Transactions of Human-Robot Interaction* 13 (2024), 1–30. DOI: https://doi.org/10.1145/3656376.

[12] A. Fischer-Janzen, T. M. Wendt, and K. Van Laerhoven. "A scoping review of gaze and eye tracking-based control methods for assistive robotic arms". In *Frontiers in Robotics and AI* 11 (2024), 1–22. DOI: https://doi.org/10.3389/frobt.2024.1326670.

[13] H. Admoni and S. Srinivasa. "Predicting User Intent Through Eye Gaze for Shared Autonomy". In *The 2016 AAAI Fall Symposium Series: Shared Autonomy in Research and Practice Technical Report FS-16-05*. 2016, 298–303.

[14] F. Koochaki and L. Najafizadeh. "A Data-Driven Framework for Intention Prediction via Eye Movement With Applications to Assistive Systems". In *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 29 (2021), 974–984. DOI: https://doi.org/10.1109/TNSRE.2021.3083815.

[15] T. Bhattacharjee, E. K. Gordon, R. Scalise, M. E. Cabrera, A. Caspi, M. Cakmak, and S. S. Srinivasa. "Is More Autonomy Always Better?" In *Proceedings of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. New York, NY, USA: ACM, 2020, 181–190. DOI: https://doi.org/10.1145/3319502.3374818.

[16] M. H. Farhadi, A. Rabiee, S. Ghafoori, A. Cetera, W. Xu, and R. Abiri. "Human-Centered Shared Autonomy for Motor Planning, Learning, and Control Applications". In. *Bridging the Gap between Mind and Machine: Exploring the Future of Human-AI-Neurotechnology Integration*. Cham: Springer Nature Switzerland, 2026, 259–297. DOI: https://doi.org/10.1007/978-3-032-06713-5_14.

[17] P. M. Tostado, W. W. Abbott, and A. A. Faisal. "3D gaze cursor: Continuous calibration and end-point grasp control of robotic actuators". In *2016 IEEE International Conference on Robotics and Automation (ICRA)*. IEEE, 2016, 3295–3300. DOI: https://doi.org/10.1109/ICRA.2016.7487502.

[18] J. M. Catalán, J. A. Díez, A. Bertomeu-Motos, F. J. Badesa, and N. Garcia-Aracil. "Multimodal Control Architecture for Assistive Robotics". In *Converging Clinical and Engineering Research on Neurorehabilitation II*. Vol. 15. Biosystems & Biorobotics. Springer International Publishing, 2017, 513–517. DOI: https://doi.org/10.1007/978-3-319-46669-9_85.

[19] Y.-S. L.-K. Cio, M. Raison, C. Leblond Menard, and S. Achiche. "Proof of Concept of an Assistive Robotic Arm Control Using Artificial Stereovision and Eye-Tracking". In *IEEE Transactions on Neural Systems and Rehabilitation Engineering* 27, 12 (2019), 2344–2352. DOI: https://doi.org/10.1109/TNSRE.2019.2950619.

[20] B. Yang, J. Huang, M. Sun, J. Huo, X. Li, and C. Xiong. "Head-free, Human Gaze-driven Assistive Robotic System for Reaching and Grasping". In *2021 40th Chinese Control Conference (CCC)*. IEEE, 2021, 4138–4143. DOI: https://doi.org/10.23919/CCC52363.2021.9549800.

[21] A. Fischer-Janzen, T. M. Wendt, D. Görlich, and K. Van Laerhoven. "Eye-Tracking-Driven Shared Control for Robotic Arms: Wizard of Oz Studies to Assess Design Choices". In *J. Hum.-Robot Interact.* (2026). Just Accepted, 1–30. DOI: https://doi.org/10.1145/3796524. URL: https://doi.org/10.1145/3796524.

[22] Z. Zhang, B. Yang, X. Chen, W. Shi, H. Wang, W. Luo, and J. Huang. "Mindeye-Omniassist: A Gaze-Driven LLM-Enhanced Assistive Robot System for Implicit Intention Recognition and Task Execution". In *2025 IEEE International Conference on Cyborg and Bionic Systems (CBS)*. 2025, 1–6. DOI: https://doi.org/10.1109/CBS65871.2025.11267731.

[23] E. Smith and M. Delargy. "Locked-in syndrome". In *BMJ* 330 (2005), 406–409.

[24] World Health Organization. "International Classification of Functioning, Disability, and Health: ICF". 2001. URL: https://icd.who.int/dev11/l-icf/en.

[25] R. M. Aronson, N. Almutlak, and H. Admoni. "Inferring Goals with Gaze during Teleoperated Manipulation". In *2021 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. IEEE, 2021, 7307–7314. DOI: https://doi.org/10.1109/IROS51168.2021.9636551.

[26] Z. Bien, D.-J. Kim, M.-J. Chung, D.-S. Kwon, and P.-H. Chang. "Development of a wheelchair-based rehabilitation robotic system (KARES II) with various human-robot interaction interfaces for the disabled". In *Proceedings 2003 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM 2003)*. IEEE, 2003, 902–907. DOI: https://doi.org/10.1109/AIM.2003.1225462.

[27] S. Alsharif, O. Kuzmicheva, and A. Gräser. "Gaze Gesture-Based Human Robot Interface". In *Technische Unterstützungssysteme, die die Menschen wirklich wollen*. 2016, 339–348. URL: https://www.researchgate.net/publication/313896309_Gaze_Gesture-Based_Human_Robot_Interface.

[28] M. Di Maio, P. Dondi, L. Lombardi, and M. Porta. "Hybrid Manual and Gaze-Based Interaction With a Robotic Arm". In *2021 26th IEEE International Conference on Emerging Technologies and Factory Automation (ETFA)*. IEEE, 2021, 1–4. DOI: https://doi.org/10.1109/ETFA45728.2021.9613371.

[29] A. Al-Rahayfeh and M. Faezipour. "Eye Tracking and Head Movement Detection: A State-of-Art Survey". In *IEEE journal of translational engineering in health and medicine* 1 (2013), 210–212. DOI: https://doi.org/10.1109/JTEHM.2013.2289879.

[30] R. E. Cowan, B. J. Fregly, M. L. Boninger, L. Chan, M. M. Rodgers, and D. J. Reinkensmeyer. "Recent trends in assistive technology for mobility". In *Journal of neuroengineering and rehabilitation* 9, 20 (2012).

[31] K. Holmqvist and R. Andersson. "Eye-tracking: A comprehensive guide to methods, paradigms and measures". Oxford University Press, 2017.

[32] Connect Tobii. "Dark and bright pupil tracking". June 1, 2023. URL: https://connect.tobii.com/s/article/What-is-dark-and-bright-pupil-tracking?language=en_US.

[33] A. F. Klaib, N. O. Alsrehin, W. Y. Melhem, H. O. Bashtawi, and A. A. Magableh. "Eye tracking algorithms, techniques, tools, and applications with an emphasis on machine learning and Internet of Things technologies". In *Expert Systems with Applications* 166 (2021), 1–24. DOI: https://doi.org/10.1016/j.eswa.2020.114037.

[34] U. Chaudhary, I. Vlachos, J. B. Zimmermann, A. Espinosa, A. Tonin, A. Jaramillo-Gonzalez, M. Khalili-Ardali, H. Topka, J. Lehmberg, G. M. Friehs, A. Woodtli, J. P. Donoghue, and N. Birbaumer. "Spelling interface using intracortical signals in a completely locked-in patient enabled via auditory neurofeedback training". In *Nature Communications* 13, 1 (2022), 1236. DOI: https://doi.org/10.1038/s41467-022-28859-8.

[35] K. Karas, L. Pozzi, A. Pedrocchi, F. Braghin, and L. Roveda. “Brain-computer interface for robot control with eye artifacts for assistive applications”. In *Scientific reports* 13, 1 (2023), 17512. DOI: https://doi.org/10.1038/s41598-023-44645-y.

[36] Q. Huang, Z. Zhang, T. Yu, S. He, and Y. Li. “An EEG-/EOG-Based Hybrid Brain-Computer Interface: Application on Controlling an Integrated Wheelchair Robotic Arm System”. In *Frontiers in neuroscience* 13 (2019), 1–9. DOI: https://doi.org/10.3389/fnins.2019.01243.

[37] A. C. Tricco, E. Lillie, W. Zarin, K. K. O’Brien, H. Colquhoun, D. Levac, D. Moher, M. D. J. Peters, T. Horsley, L. Weeks, S. Hempel, E. A. Akl, C. Chang, J. McGowan, L. Stewart, L. Hartling, A. Aldcroft, M. G. Wilson, C. Garritty, S. Lewin, C. M. Godfrey, M. T. Macdonald, E. V. Langlois, K. Soares-Weiser, J. Moriarty, T. Clifford, Ö. Tunçalp, and S. E. Straus. “PRISMA Extension for Scoping Reviews (PRISMA-ScR): Checklist and Explanation”. In *Annals of internal medicine* 169, 7 (2018), 467–473. DOI: https://doi.org/10.7326/M18-0850.

[38] A. Fischer-Janzen. “AnkeLinus/EyeTrackingInRobotControlTasks: V4 Updated Version of data base ”. 2023. DOI: https://doi.org/10.5281/zenodo.10034509. URL: https://github.com/AnkeLinus/EyeTrackingInRobotControlTasks/tree/main.

[39] D. H. Kim, J. H. Kim, D. H. Yoo, Y. J. Lee, and M. J. Chung. “A Human-Robot Interface Using Eye-Gaze Tracking System for People with Motor Disabilities”. In *Transaction on Control, Automation and Systems Engineering* 3, 4 (2001), 229–235.

[40] D. H. Yoo, J. H. Kim, D. H. Kim, and M. J. Chung. “A human-robot interface using vision-based eye gaze estimation system”. In *IEEE/RSJ International Conference on Intelligent Robots and System*. IEEE, 2002, 1196–1201. DOI: https://doi.org/10.1109/IRDS.2002.1043896.

[41] Z. Bien, M.-J. Chung, P.-H. Chang, D.-S. Kwon, D.-J. Kim, J.-S. Han, J.-H. Kim, D.-H. Kim, H.-S. Park, S.-H. Kang, K. Lee, and S.-C. Lim. “Integration of a Rehabilitation Robotic System (KARES II) with Human-Friendly Man-Machine Interaction Units”. In *Autonomous Robots* 16, 2 (2004), 165–191. DOI: https://doi.org/10.1023/B:AURO.0000016864.12513.77.

[42] E. Iáñez, J. M. Azorín, E. Fernández, and A. Úbeda. “Interface based on electrooculography for velocity control of a robot arm”. In *Applied Bionics and Biomechanics* 7, 3 (2010), 199–207. DOI: https://doi.org/10.1080/11762322.2010.503107.

[43] D. P. McMullen, G. Hotson, K. D. Katyal, B. A. Wester, M. S. Fifer, T. G. McGee, A. Harris, M. S. Johannes, R. J. Vogelstein, A. D. Ravitz, W. S. Anderson, N. V. Thakor, and N. E. Crone. “Demonstration of a semi-autonomous hybrid brain-machine interface using human intracranial EEG, eye tracking, and computer vision to control a robotic upper limb prosthetic”. In *IEEE transactions on neural systems and rehabilitation engineering* 22, 4 (2014), 784–796. DOI: https://doi.org/10.1109/TNSRE.2013.2294685.

[44] J. D. Webb, S. Li, and X. Zhang. “Using visuomotor tendencies to increase control performance in teleoperation”. In *2016 American Control Conference (ACC)*. IEEE, 2016, 7110–7116. DOI: https://doi.org/10.1109/ACC.2016.7526794.

[45] H. Zeng, Y. Wang, C. Wu, A. Song, J. Liu, P. Ji, B. Xu, L. Zhu, H. Li, and P. Wen. “Closed-Loop Hybrid Gaze Brain-Machine Interface Based Robotic Arm Control with Augmented Reality Feedback”. In *Frontiers in neurorobotics* 11, 60 (2017), 1–13. DOI: https://doi.org/10.3389/fnbot.2017.00060.

[46] E. Jones, W. Chinthammit, W. Huang, U. Engelke, and C. Lueg. “Symmetric Evaluation of Multimodal Human–Robot Interaction with Gaze and Standard Control”. In *Symmetry* 10, 12 (2018), 1–15. DOI: https://doi.org/10.3390/sym10120680.

[47] V. K. Sharma, K. Saluja, V. Mollyn, and P. Biswas. "Eye Gaze Controlled Robotic Arm for Persons with Severe Speech and Motor Impairment". In *ACM Symposium on Eye Tracking Research and Applications*. New York, NY, USA: ACM, 2020, 1–9. DOI: https://doi.org/10.1145/3379155.3391324.

[48] S. Stalljann, L. Wöhle, J. Schäfer, and M. Gebhard. "Performance Analysis of a Head and Eye Motion-Based Control Interface for Assistive Robots". In *Sensors (Basel, Switzerland)* 20, 24 (2020). DOI: https://doi.org/10.3390/s20247162.

[49] L. Scalera, S. Seriani, P. Gallina, M. Lentini, and A. Gasparetto. "Human-Robot Interaction through Eye Tracking for Artistic Drawing". In *Robotics* 10, 2 (2021), 54. DOI: https://doi.org/10.3390/robotics10020054.

[50] L. Scalera, S. Seriani, A. Gasparetto, and P. Gallina. "A Novel Robotic System for Painting with Eyes". In *Advances in Italian Mechanism Science*. Vol. 91. Mechanisms and Machine Science. Cham: Springer International Publishing, 2021, 191–199. DOI: https://doi.org/10.1007/978-3-030-55807-9_22.

[51] M. S. H. Sunny, M. I. I. Zarif, I. Rulik, J. Sanjuan, M. H. Rahman, S. I. Ahamed, I. Wang, K. Schultz, and B. Brahmi. "Eye-Gaze Control of a Wheelchair Mounted 6DOF Assistive Robot for Activities of Daily Living". In *Journal of NeuroEngineering and Rehabilitation* 18 (2021). DOI: https://doi.org/10.1186/s12984-021-00969-2.

[52] A. Dragomir, C. F. Pana, D. Cojocaru, and L. F. Manga. "Human-Machine Interface for Controlling a Light Robotic Arm by Persons with Special Needs". In *2021 22nd International Carpathian Control Conference (ICCC)*. Velké Karlovice, Czech Republic: IEEE, 2021, 1–6. DOI: https://doi.org/10.1109/ICCC51557.2021.9454664.

[53] V. K. Sharma, L. R. D. Murthy, and P. Biswas. "Comparing Two Safe Distance Maintenance Algorithms for a Gaze-Controlled HRI Involving Users with SSMI". In *ACM Transactions on Accessible Computing* 15, 3 (2022), 1–23. DOI: https://doi.org/10.1145/3530822.

[54] A. Bannat, J. Gast, T. Rehrl, W. Rösel, G. Rigoll, and F. Wallhoff. "A Multimodal Human-Robot-Interaction Scenario: Working Together with an Industrial Robot". In *Human-Computer Interaction. Novel Interaction Methods and Techniques*. Vol. 5611. Lecture Notes in Computer Science. Berlin, Heidelberg: Springer Berlin Heidelberg, 2009, 302–311. DOI: https://doi.org/10.1007/978-3-642-02577-8_33.

[55] A. Ubeda, E. Iañez, and J. M. Azorin. "Wireless and Portable EOG-Based Interface for Assisting Disabled People". In *IEEE/ASME Transactions on Mechatronics* 16, 5 (2011), 870–873. DOI: https://doi.org/10.1109/TMECH.2011.2160354.

[56] A. Khan, M. A. Memon, Y. Jat, and A. Khan. "Electro-Occulogram Based Interactive Robotic Arm Interface for Partially Paralytic Patients". In *ITEE Journal* 1, 1 (2012).

[57] J. Zhang, F. Guo, J. Hong, and Y. Zhang. "Human-robot shared control of articulated manipulator". In *2013 IEEE International Symposium on Assembly and Manufacturing (ISAM)*. IEEE, 2013, 81–84. DOI: https://doi.org/10.1109/ISAM.2013.6643493.

[58] H. Wang, X. Dong, Z. Chen, and B. E. Shi. "Hybrid gaze/EEG brain computer interface for robot arm control on a pick and place task". In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference* 2015 (2015), 1476–1479. DOI: https://doi.org/10.1109/EMBC.2015.7318649.

[59] S. Dziemian, W. W. Abbott, and A. A. Faisal. "Gaze-based teleprosthetic enables intuitive continuous control of complex robot arm use: Writing & drawing". In *2016 6th IEEE International Conference on Biomedical Robotics and Biomechatronics (BioRob)*. IEEE, 2016, 1277–1282. DOI: https://doi.org/10.1109/BIOROB.2016.7523807.

[60] Y. Wang, G. Xu, A. Song, B. Xu, H. Li, C. Hu, and H. Zeng. "Continuous Shared Control for Robotic Arm Reaching Driven by a Hybrid Gaze-Brain Machine Interface". In *2018 IEEE/RSJ International Conference on Intelligent Robots and Systems*. Madrid, Spain: IEEE, 2018. DOI: https://doi.org/10.1109/IROS.2018.8594367.

[61] F. D. Perez Reynoso, P. A. Niño Suarez, O. F. Aviles Sanchez, M. B. Calva Yañez, E. Vega Alvarado, and E. A. Portilla Flores. "A Custom EOG-Based HMI Using Neural Network Modeling to Real-Time for the Trajectory Tracking of a Manipulator Robot". In *Frontiers in neurorobotics* 14 (2020), 1–23. DOI: https://doi.org/10.3389/fnbot.2020.578834.

[62] M. Rusydi, T. Okamoto, S. Ito, and M. Sasaki. "Rotation Matrix to Operate a Robot Manipulator for 2D Analog Tracking Objects Using Electrooculography". In *Robotics* 3, 3 (2014), 289–309. DOI: https://doi.org/10.3390/robotics3030289.

[63] M. I. Rusydi, M. Sasaki, and S. Ito. "Affine transform to reform pixel coordinates of EOG signals for controlling robot manipulators using gaze motions". In *Sensors (Basel, Switzerland)* 14, 6 (2014), 10107–10123. DOI: https://doi.org/10.3390/s140610107.

[64] G. Onose, C. Grozea, A. Anghelescu, C. Daia, C. J. Sinescu, A. V. Ciurea, T. Spircu, A. Mirea, I. Andone, A. Spânu, C. Popescu, A.-S. Mihăescu, S. Fazli, M. Danóczy, and F. Popescu. "On the feasibility of using motor imagery EEG-based brain-computer interface in chronic tetraplegics for assistive robotic arm control: a clinical test and long-term post-trial follow-up". In *Spinal Cord* 50, 8 (2012), 599–608. DOI: https://doi.org/10.1038/sc.2012.14.

[65] C.-M. Huang and B. Mutlu. "Anticipatory robot control for efficient human-robot collaboration". In *2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*. IEEE, 2016, 83–90. DOI: https://doi.org/10.1109/HRI.2016.7451737.

[66] S. Li, X. Zhang, and J. D. Webb. "3-D-Gaze-Based Robotic Grasping Through Mimicking Human Visuomotor Function for People With Motion Impairments". In *IEEE transactions on bio-medical engineering* 64, 12 (2017), 2824–2835. DOI: https://doi.org/10.1109/TBME.2017.2677902.

[67] E. Ivorra, M. Ortega, J. M. Catalán, S. Ezquerro, L. D. Lledó, N. Garcia-Aracil, and M. Alcañiz. "Intelligent Multimodal Framework for Human Assistive Robotics Based on Computer Vision Algorithms". In *Sensors (Basel, Switzerland)* 18, 8 (2018). DOI: https://doi.org/10.3390/s18082408.

[68] L. Wöhle and M. Gebhard. "Towards Robust Robot Control in Cartesian Space Using an Infrastructureless Head- and Eye-Gaze Interface". In *Sensors* 21, 5 (2021). DOI: https://doi.org/10.3390/s21051798.

[69] K.-B. Park, S. H. Choi, H. Moon, J. Y. Lee, Y. Ghasemi, and H. Jeong. "Indirect Robot Manipulation using Eye Gazing and Head Movement For Future of Work in Mixed Reality". In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 2022, 483–484. DOI: https://doi.org/10.1109/VRW55335.2022.00107.

[70] S. Alsharif. "Gaze-Based Control of Robot Arm in Three-Dimensional Space". PhD thesis. Bremen: University Bremen, November 27, 2017. URL: https://media.suub.uni-bremen.de/bitstream/elib/1394/1/00106451-1.pdf.

[71] M. I. Shahzad and S. Mehmood. "Control of Articulated Robot Arm by Eye Tracking". Master Thesis. SE – 37179 Karlsakrona: School of Computing Blekinge Institute of Technology, September 2010. URL: https://www.diva-portal.org/smash/get/diva2:830394/FULLTEXT01.pdf.

[72] J. Nunez-Varela. "Gaze control for visually guided manipulation". PhD Thesis. Birmingham: University of Birmingham, December 2012. URL: https://etheses.bham.ac.uk/id/eprint/4444/1/Nunez-Varela13PhD1.pdf.

[73] M. J. Payton David W.; Daily. "Systems, Methods, and Apparatus for Neuro-Robotic tracking point selection". US8483816 B1, 2013.

[74] E. R. Norales, D. Brasset, and P. Invernizzi. "Robotized Surgery System with improved Control". US9360934 B2, 2016.

[75] C. Mejia and Y. Kajikawa. "Bibliometric Analysis of Social Robotics Research: Identifying Research Trends and Knowledgebase". In *Applied Sciences* 7, 12 (2017), 1316. DOI: https://doi.org/10.3390/app7121316.

[76] G. Zhang, J. P. Hansen, K. Minakata, A. Alapetite, and Z. Wang. "Eye-Gaze-Controlled Telepresence Robots for People with Motor Disabilities". In *The 14th ACM/IEEE International Conference on Human-Robot Interaction*. Daegu, South Korea: IEEE, 2019.

[77] M. Callejas-Cuervo, A. X. González-Cely, and T. Bastos-Filho. "Control Systems and Electronic Instrumentation Applied to Autonomy in Wheelchair Mobility: The State of the Art". In *Sensors (Basel, Switzerland)* 20, 21 (2020). DOI: https://doi.org/10.3390/s20216326.

[78] H. Jiang, T. Zhang, J. P. Wachs, and B. S. Duerstock. "Enhanced control of a wheelchair-mounted robotic manipulator using 3-D vision and multimodal interaction". In *Computer Vision and Image Understanding* 149 (2016), 21–31. DOI: https://doi.org/10.1016/j.cviu.2016.03.015.

[79] C. McMurrough, S. Ferdous, A. Papangelis, A. Boisselle, and F. Makedon. "A Survey of Assistive Devices for Cerebral Palsy Patients". In *Proceedings of the 5th International Conference on PErvasive Technologies Related to Assistive Environments*. ACM, 2012, 1–8. DOI: https://doi.org/10.1145/2413097.

[80] C. McMurrough, I. Ranatunga, A. Papangelis, D. O. Popa, and F. Makedon. "A development and evaluation platform for non-tactile power wheelchair controls". In *Proceedings of the 6th International Conference on PErvasive Technologies Related to Assistive Environments - PETRA '13*. New York, New York, USA: ACM Press, 2013, 1–4. DOI: https://doi.org/10.1145/2504335.2504339.

[81] Y. K. Meena, A. Chowdhury, H. Cecotti, K. Wong-Lin, S. S. Nishad, A. Dutta, and G. Prasad. "EMOHEX: An eye tracker based mobility and hand exoskeleton device for assisting disabled people". In *2016 IEEE International Conference on Systems, Man, and Cybernetics (SMC)*. IEEE, 2016, 002122–002127. DOI: https://doi.org/10.1109/SMC.2016.7844553.

[82] Y. K. Meena, H. Cecotti, K. Wong-Lin, and G. Prasad. "A multimodal interface to resolve the Midas-Touch problem in gaze controlled wheelchair". In *Annual International Conference of the IEEE Engineering in Medicine and Biology Society. IEEE Engineering in Medicine and Biology Society. Annual International Conference* 2017 (2017), 905–908. DOI: https://doi.org/10.1109/EMBC.2017.8036971.

[83] J. Ma, Y. Zhang, A. Cichocki, and F. Matsuno. "A novel EOG/EEG hybrid human-machine interface adopting eye movements and ERPs: application to robot control". In *IEEE transactions on bio-medical engineering* 62, 3 (2015), 876–889. DOI: https://doi.org/10.1109/TBME.2014.2369483.

[84] N. Robinson, B. Tidd, D. Campbell, D. Kulić, and P. Corke. "Robotic Vision for Human-Robot Interaction and Collaboration: A Survey and Systematic Review". In *ACM Transactions on Human-Robot Interaction* (2022). DOI: https://doi.org/10.1145/3570731.

[85] J. A. Gonzalez-Aguirre, R. Osorio-Oliveros, K. L. Rodríguez-Hernández, J. Lizárraga-Iturralde, R. Morales Menendez, R. A. Ramírez-Mendoza, M. A. Ramírez-Moreno, and J. d. J. Lozoya-Santos. "Service Robots: Trends and Technology". In *Applied Sciences* 11, 22 (2021), 10702. DOI: https://doi.org/10.3390/app112210702.

[86] N. Liang and G. Nejat. "A Meta-Analysis on Remote HRI and In-Person HRI: What Is a Socially Assistive Robot to Do?" In *Sensors (Basel, Switzerland)* 22, 19 (2022). DOI: https://doi.org/10.3390/s22197155.

[87] O. Nocentini, L. Fiorini, G. Acerbi, A. Sorrentino, G. Mancioppi, and F. Cavallo. "A Survey of Behavioral Models for Social Robots". In *Robotics* 8, 3 (2019), 54. DOI: https://doi.org/10.3390/robotics8030054.

[88] F. Despinoy, L. Torres, J. Roberto, M.-A. Vitrani, and B. Herman. "Toward remote teleoperation with eye and hand: A first experimental study". In *3rd Joint Workshop on New Technologies for Computer/Robot Assisted Surgery (CRAS2013)*. 2013, 1–4.

[89] S. Jain and B. Argall. "Probabilistic Human Intent Recognition for Shared Autonomy in Assistive Robotics". In *ACM transactions on human-robot interaction* 9, 1 (2019). DOI: https://doi.org/10.1145/3359614.

[90] R. M. Aronson, T. Santini, T. C. Kübler, E. Kasneci, S. Srinivasa, and H. Admoni. "Eye-Hand Behavior in Human-Robot Shared Manipulation". In *Proceedings of the 2018 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2018, 4–13. DOI: https://doi.org/10.1145/3171221.3171287.

[91] R. M. Aronson and H. Admoni. "Semantic gaze labeling for human-robot shared manipulation". In *Proceedings of the 11th ACM Symposium on Eye Tracking Research & Applications*. ACM, 2019, 1–9. DOI: https://doi.org/10.1145/3314111.3319840.

[92] R. M. Aronson and H. Admoni. "Eye Gaze for Assistive Manipulation". In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. ACM, 2020, 552–554. DOI: https://doi.org/10.1145/3371382.3377434.

[93] C.-C. Postelnicu, F. Girbacia, G.-D. Voinea, and R. Boboc. "Towards Hybrid Multimodal Brain Computer Interface for Robotic Arm Command". In *Augmented cognition*. Cham: Springer, 2019, 461–470.

[94] F. Cincotti, D. Mattia, F. Aloise, S. Bufalari, G. Schalk, G. Oriolo, A. Cherubini, M. G. Marciani, and F. Babiloni. "Non-invasive brain-computer interface system: towards its application as assistive technology". In *Brain research bulletin* 75, 6 (2008), 796–803. DOI: https://doi.org/10.1016/j.brainresbull.2008.01.007.

[95] S. Crea, M. Nann, E. Trigili, F. Cordella, A. Baldoni, F. J. Badesa, J. M. Catalán, L. Zollo, N. Vitiello, N. G. Aracil, and S. R. Soekadar. "Feasibility and safety of shared EEG/EOG and vision-guided autonomous whole-arm exoskeleton control to perform activities of daily living". In *Scientific reports* 8, 1 (2018), 10823. DOI: https://doi.org/10.1038/s41598-018-29091-5.

[96] E. A. Kirchner, J. C. Albiez, A. Seeland, M. Jordan, and F. Kirchner. "Towards Assistive Robotics for Home Rehabilitation". In *Proceedings of the International Conference on Biomedical Electronics and Devices*. SciTePress - Science and and Technology Publications, 2013, 168–177. DOI: https://doi.org/10.5220/0004248501680177.

[97] A. Shafti, P. Orlov, and A. A. Faisal. “Gaze-based, Context-aware Robotic System for Assisted Reaching and Grasping”. In *2019 International Conference on Robotics and Automation (ICRA)*. 2019, 863–869. DOI: https://doi.org/10.1109/ICRA.2019.8793804.

[98] A. Shafti and A. A. Faisal. “Non-invasive Cognitive-level Human Interfacing for the Robotic Restoration of Reaching & Grasping”. In *2021 10th International IEEE/EMBS Conference on Neural Engineering (NER)*. IEEE, 2021, 872–875. DOI: https://doi.org/10.1109/NER49283.2021.9441453.

[99] R. O. Maimon-Dror, J. Fernandez-Quesada, G. A. Zito, C. Konnaris, S. Dziemian, and A. A. Faisal. “Towards free 3D end-point control for robotic-assisted human reaching using binocular eye tracking”. In *IEEE International Conference on Rehabilitation Robotics* 2017 (2017), 1049–1054. DOI: https://doi.org/10.1109/ICORR.2017.8009388.

[100] A. Pedrocchi, S. Ferrante, E. Ambrosini, M. Gandolla, C. Casellato, T. Schauer, C. Klauer, J. Pascual, C. Vidaurre, M. Gföhler, W. Reichenfelser, J. Karner, S. Micera, A. Crema, F. Molteni, M. Rossini, G. Palumbo, E. Guanziroli, A. Jedlitschka, M. Hack, M. Bulgheroni, E. d’Amico, P. Schenk, S. Zwicker, A. Duschau-Wicke, J. Miseikis, L. Graber, and G. Ferrigno. “MUNDUS project: MUltimodal neuroprosthesis for daily upper limb support”. In *Journal of neuroengineering and rehabilitation* 10 (2013), 1–20. DOI: https://doi.org/10.1186/1743-0003-10-66.

[101] B. Noronha, S. Dziemian, G. A. Zito, C. Konnaris, and A. A. Faisal. “"Wink to grasp" - comparing eye, voice & EMG gesture control of grasp with soft-robotic gloves”. In *IEEE International Conference on Rehabilitation Robotics* 2017 (2017), 1043–1048. DOI: https://doi.org/10.1109/ICORR.2017.8009387.

[102] D. Novak and R. Riener. “Enhancing patient freedom in rehabilitation robotics using gaze-based intention detection”. In *IEEE International Conference on Rehabilitation Robotics* 2013 (2013), 6650507. DOI: https://doi.org/10.1109/ICORR.2013.6650507.

[103] I. S. Shehu, Y. Wang, A. M. Athuman, and X. Fu. “Remote Eye Gaze Tracking Research: A Comparative Evaluation on Past and Recent Progress”. In *Electronics* 10, 24 (2021), 3165. DOI: https://doi.org/10.3390/electronics10243165.

[104] A. Clark and I. Ahmad. “Interfacing with Robots without the use of Touch or Speech”. In *The 14th PErvasive Technologies Related to Assistive Environments Conference*. New York, NY, USA: ACM, 2021, 347–353. DOI: https://doi.org/10.1145/3453892.3461330.

[105] D. Esposito, J. Centracchio, E. Andreozzi, G. D. Gargiulo, G. R. Naik, and P. Bifulco. “Biosignal-Based Human-Machine Interfaces for Assistance and Rehabilitation: A Survey”. In *Sensors (Basel, Switzerland)* 21, 20 (2021). DOI: https://doi.org/10.3390/s21206863.

[106] S. Ramkumar, K. Sathesh Kumar, T. Dhiliphan Rajkumar, M. Ilayaraja, and K. Sankar. “A review-classification of electrooculogram based human computer interfaces”. In *Biomedical Research* 29, 6 (2018), 078–1084.

[107] J. Schäfer and M. 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*. New York, NY, USA: ACM, 2019, 482–488. DOI: https://doi.org/10.1145/3316782.3322774.

[108] A. Dünser, M. Lochner, U. Engelke, and D. Rozado Fernández. “Visual and manual control for human-robot teleoperation”. In *IEEE computer graphics and applications* 35, 3 (2015), 22–32. DOI: https://doi.org/10.1109/MCG.2015.4.

[109] E. Pasqualotto, T. Matuz, S. Federici, C. A. Ruf, M. Bartl, M. Olivetti Belardinelli, N. Birbaumer, and S. Halder. "Usability and Workload of Access Technology for People With Severe Motor Impairment: A Comparison of Brain-Computer Interfacing and Eye Tracking". In *Neurorehabilitation and neural repair* 29, 10 (2015), 950–957. DOI: https://doi.org/10.1177/1545968315575611.

[110] A.-I. Siean and R.-D. Vatavu. "Wearable Interactions for Users with Motor Impairments: Systematic Review, Inventory, and Research Implications". In *The 23rd International ACM SIGACCESS Conference on Computers and Accessibility*. New York, NY, USA: ACM, 2021, 1–15. DOI: https://doi.org/10.1145/3441852.3471212.

[111] H. Drewes and A. Schmidt. "Interacting with the Computer Using Gaze Gestures". In *Human-Computer Interaction - INTERACT 2007*. Springer, Berlin, Heidelberg, 2007, 1–14.

[112] B. B. Velichkovsky, M. A. Rumyantsev, and M. A. Morozov. "New Solution to the Midas Touch Problem: Identification of Visual Commands Via Extraction of Focal Fixations". In *Procedia Computer Science* 39 (2014), 75–82. DOI: https://doi.org/10.1016/j.procs.2014.11.012.

[113] R. Bednarik, T. Gowases, and M. Tukiainen. "Gaze interaction enhances problem solving: Effects of dwell-time based, gaze-augmented, and mouse interaction on problem-solving strategies and user experience". In *Journal of Eye Movement Research* 3, 1 (2009). DOI: https://doi.org/10.16910/jemr.3.1.3.

[114] M. Duguleana and G. Mogan. "Using Eye Blinking for EOG-Based Robot Control". In *Emerging trends in technological innovation. DoCEIS 2010*. Springer, 2010, 343–350.

[115] A. Jelinek. "Guidelines 3/2019 on processing of personal data through video devices". 2019. URL: https://edpb.europa.eu/sites/default/files/consultation%20/edpb_guidelines_201903_videosurveillance.pdf.

[116] World Health Organization. "World report on vision". 2019. URL: https://www.who.int/publications/i/item/9789241516570.

[117] K.-S. Hong and M. J. Khan. "Hybrid Brain-Computer Interface Techniques for Improved Classification Accuracy and Increased Number of Commands: A Review". In *Frontiers in neurorobotics* 11 (2017), 35. DOI: https://doi.org/10.3389/fnbot.2017.00035.

[118] L. F. Nicolas-Alonso and J. Gomez-Gil. "Brain computer interfaces, a review". In *Sensors (Basel, Switzerland)* 12, 2 (2012), 1211–1279. DOI: https://doi.org/10.3390/s120201211.

[119] L. R. Trambaiolli and T. H. Falk. "Hybrid brain–computer interfaces for wheelchair control: a review of existing solutions, their advantages and open challenges". In *Smart Wheelchairs and Brain-Computer Interfaces*. Elsevier, 2018, 229–256. DOI: https://doi.org/10.1016/B978-0-12-812892-3.00010-8.

[120] A. Bonci, P. D. Cen Cheng, M. Indri, G. Nabissi, and F. Sibona. "Human-Robot Perception in Industrial Environments: A Survey". In *Sensors (Basel, Switzerland)* 21, 5 (2021). DOI: https://doi.org/10.3390/s21051571.

[121] A. Hentout, M. Aouache, A. Maoudj, and I. Akli. "Human–robot interaction in industrial collaborative robotics: a literature review of the decade 2008–2017". In *Advanced Robotics* 33, 15-16 (2019), 764–799. DOI: https://doi.org/10.1080/01691864.2019.1636714.

[122] S. G. Hart and L. E. Staveland. "Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research". In *Human Mental Workload*. Vol. 52. Advances in Psychology. Elsevier, 1988, 139–183. DOI: https://doi.org/10.1016/S0166-4115(08)62386-9.

[123] R. D. McKendrick and E. Cherry. "A Deeper Look at the NASA TLX and Where It Falls Short". In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 62, 1 (2018), 44–48. DOI: https://doi.org/10.1177/1541931218621010.

[124] A. Lund. "Measuring Usability with the USE Questionnaire". In *Usability and User Experience Newsletter of the STC Usability SIG* 8 (2001).

[125] M. Gao, P. Kortum, and F. Oswald. "Psychometric Evaluation of the USE (Usefulness, Satisfaction, and Ease of use) Questionnaire for Reliability and Validity". In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 62, 1 (2018), 1414–1418. DOI: https://doi.org/10.1177/1541931218621322.

[126] J. Brooke. "SUS: A 'Quick and Dirty' Usability Scale". In *Usability Evaluation in Industry*. London: CRC Press, 1996, 1–7.

[127] L. Demers, R. Weiss-Lambrou, and B. Ska. "Development of the Quebec User Evaluation of Satisfaction with assistive Technology (QUEST)". In *Assistive technology* 8, 1 (1996), 3–13. DOI: https://doi.org/10.1080/10400435.1996.10132268.

[128] L. Demers, M. Monette, Y. Lapierre, D. Arnold, and C. Wolfson. "Reliability, validity, and applicability of the Quebec User Evaluation of Satisfaction with assistive Technology (QUEST 2.0) for adults with multiple sclerosis". In *Disability and Rehabilitation* 24, 1-3 (2002). DOI: https://doi.org/10.1080/09638280110066352.

[129] J. Schmidtler, K. Bengler, F. Dimeas, and A. Campeau-Lecours. "Questionnaire for the Evaluation of Physical Assistive Devices (QUEAD) - Manual". In *ResearchGate* (2017). DOI: https://doi.org/10.13140/RG.2.2.31113.44649.

[130] D. Lulé, S. Häcker, A. Ludolph, N. Birbaumer, and A. Kübler. "Depression und Lebensqualität bei Patienten mit amyotropher Lateralsklerose". In *Deutsches Arzteblatt international* 105 (2008), 397–403. DOI: https://doi.org/10.3238/arztebl.2008.0397.

[131] C.-S. Chung, H. Wang, and R. A. Cooper. "Functional assessment and performance evaluation for assistive robotic manipulators: Literature review". In *The journal of spinal cord medicine* 36, 4 (2013), 273–289. DOI: https://doi.org/10.1179/2045772313Y.0000000132.

[132] A. Tajadura-Jiménez, A. Väljamäe, F. Bevilacqua, and N. Bianchi-Berthouze. "Principles for Designing Body-Centered Auditory Feedback". In *The Wiley Handbook of Human Computer Interaction*. Chichester, UK: John Wiley & Sons, Ltd, 2018, 371–403. DOI: https://doi.org/10.1002/9781118976005.ch18.

[133] A. Fischer, P. Gawron, L. Stiglmeier, T. M. Wendt, and K. van Laehrhoven. "Evaluation of precision, accuracy and threshold for the design of vibrotactile feedback in eye tracking applications". In *J. Sens. Sens. Syst.* 12 (2023), 103–109.

[134] J. Rantala, J. Kangas, D. Akkil, P. Isokoski, and R. Raisamo. "Glasses with haptic feedback of gaze gestures". In *CHI '14 Extended Abstracts on Human Factors in Computing Systems*. New York, NY, USA: ACM, 2014, 1597–1602. DOI: https://doi.org/10.1145/2559206.2581163.

[135] J. Rantala, P. Majaranta, J. Kangas, P. Isokoski, D. Akkil, O. Špakov, and R. Raisamo. "Gaze Interaction With Vibrotactile Feedback: Review and Design Guidelines". In *Human–Computer Interaction* 35, 1 (2020), 1–39. DOI: https://doi.org/10.1080/07370024.2017.1306444.

[136] F. Gemperle, N. Ota, and D. Siewiorek. "Design of a wearable tactile display". In *Proceedings Fifth International Symposium on Wearable Computers*. IEEE Comput. Soc, 2001, 5–12. DOI: https://doi.org/10.1109/ISWC.2001.962082.

[137] S. Choi and K. J. Kuchenbecker. “Vibrotactile Display: Perception, Technology, and Applications”. In *Proceedings of the IEEE* 101, 9 (2013), 2093–2104. DOI: https://doi.org/10.1109/JPROC.2012.2221071.

[138] C. Breazeal. “Toward sociable robots”. In *Robotics and Autonomous Systems* 42 (2003), 167–175. DOI: https://doi.org/10.1016/S0921-8890(02)00373-1.

[139] N. S. Jecker. “You’ve got a friend in me: sociable robots for older adults in an age of global pandemics”. In *Ethics and Information Technology* 23 (2021). DOI: https://doi.org/10.1007/s10676-020-09546-y.

[140] S. B. P. Samarakoon, M. A. V. J. Muthugala, and A. G. P. Jayasekara. “A Review on Human-Robot Proxemics”. In *Electronics* 11 (2022), 1–21. DOI: https://doi.org/10.3390/electronics11162490.

[141] M. Ahmad, O. Mubin, and J. Orlando. “A Systematic Review of Adaptivity in Human-Robot Interaction”. In *Multimodal Technologies and Interaction* 1, 14 (2017). DOI: https://doi.org/10.3390/mti1030014.

[142] International Federation of Robotics. “World Robotics 2023”. 2023. URL: https://ifr.org/img/worldrobotics/2023W_R_extended_version.pdf.

[143] International Organization for Standardization. “ISO 8373: 2012 (en); Robots and Robotic Devices-Vocabulary”. Geneva, Switzerland, 2012.

[144] I. Lee. “Service Robots: A Systematic Literature Review”. In *Electronics* 10, 2658 (2021). DOI: https://doi.org/10.3390/electronics10212658.

[145] Lebenshilfe e.V. “Informieren: Beiträge der Lebenshilfe”. URL: https://www.lebenshilfe-kinzig-elztal.de/.

[146] WfB Haslach gGmbH. “Informieren: Beiträge der Lebenshilfe Der Trägerverein Lebenshilfe”. URL: https://www.wfb-werbeartikel.de/de/ueber-uns/Traegerverein-Lebenshilfe.

[147] Ministerium für Soziales, Gesundheit und Integration Baden-Württemberg. “Wohnen nach Maß.” URL: https://sozialministerium.baden-wuerttemberg.de/de/soziales/menschen-mit-behinderungen/wohnen.

[148] Nippon. “Robot Runway: Pepper’s Fashion Show Debut”. URL: https://www.nippon.com/en/views/b00911/.

[149] Beuth. “DIN ISO/TS 15066:2017-04, DIN SPEC 5306:2017-04 (DE): Robots and robotic devices - Collaborative robots (ISO/TS 15066:2016)”. 2017. DOI: https://doi.org/10.31030/2584636.

[150] J. Shimaya, Y. Yoshikawa, H. Kumazaki, Y. Matsumoto, M. Miyao, and H. Ishiguro. “Communication Support via a Tele-Operated Robot for Easier Talking”. In *International Journal of Social Robotics* 11 (2019). DOI: https://doi.org/10.1007/s12369-018-0497-0.

[151] A. Hireche, A. N. Belkacem, S. Jamil, and C. Chen. “NewsGPT: ChatGPT Integration for Robot-Reporter”. 2023. arXiv: 2311.06640 [cs.RO]. URL: https://arxiv.org/abs/2311.06640.

[152] T. B. Sheridan. “A review of recent research in social robotics”. In *Current Opinion in Psychology* 36 (2020), 7–12. DOI: https://doi.org/10.1016/j.copsyc.2020.01.003.

[153] I. Papadopoulos, C. Koulouglioti, R. Lazzarino, and S. Ali. “Enablers and barriers to the implementation of socially assistive humanoid robots in health and social care: a systematic review”. In *BMJ Open* 10, 1 (2020). DOI: https://doi.org/10.1136/bmjopen-2019-033096.

[154] D. Karunarathne, Y. Morales, T. Nomura, T. Kanda, and H. Ishiguro. "Will Older Adults Accept a Humanoid Robot as a Walking Partner?" In *International Journal of Social Robotics* 11 (2019). DOI: https://doi.org/10.1007/s12369-018-0503-6.

[155] E. Fosch-Villaronga, C. Lutz, and A. Tamò-Larrieux. "Gathering Expert Opinions for Social Robots' Ethical, Legal, and Societal Concerns: Findings from Four International Workshops". In *International Journal of Social Robotics* 12 (2020). DOI: https://doi.org/10.1007/s12369-019-00605-z.

[156] W. A. Awuah, A. Ahluwalia, K. Darko, V. Sanker, J. K. Tan, P. O. Tenkorang, A. Ben-Jaafar, S. Ranganathan, N. Aderinto, A. Mehta, M. H. Shah, K. Lee Boon Chun, T. Abdul-Rahman, and O. Atallah. "Bridging Minds and Machines: The Recent Advances of Brain-Computer Interfaces in Neurological and Neurosurgical Applications". In *World Neurosurgery* 189 (2024), 138–153. DOI: https://doi.org/10.1016/j.wneu.2024.05.104.

[157] A. Baumeister, E. Gardo, P. Tolle, B. Klein, M. Pascher, J. Gerken, F. 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*. Frankfurt am Main, Germany: Deutsches Forschungszentrum für Künstliche Intelligenz GmbH (DFKI), 2021, 1–9. DOI: https://doi.org/10.48718/8p7x-cw14.

[158] I. González-Díaz, M. Molina-Moreno, J. Benois-Pineau, and A. de Rugy. "Asymmetric Multi-Task Learning for Interpretable Gaze-Driven Grasping Action Forecasting". In *IEEE Journal of Biomedical and Health Informatics* 28, 12 (2024), 7517–7530. DOI: https://doi.org/10.1109/JBHI.2024.3430810.

[159] A. Mohebbi. "Human-Robot Interaction in Rehabilitation and Assistance: a Review". In *Current Robotics Reports* 1 (2020), 131–144. DOI: https://doi.org/10.1007/s43154-020-00015-4.

[160] K. Kronhardt, M. Pascher, and J. Gerken. "Understanding Shared Control for Assistive Robotic Arms". In *VAT-HRI '23: Workshop on Variable Autonomy for Human-Robot Teaming at the ACM/IEEE International Conference on Human-Robot Interaction*. 2023. URL: http://arxiv.org/abs/2303.01993.

[161] 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*. Proc ACM SIGCHI, 2016, 35–42. DOI: https://doi.org/10.1109/HRI.2016.7451731.

[162] A. Nanavati, M. Pascher, V. Ranganeni, E. K. Gordon, T. K. Faulkner, S. S. Srinivasa, M. Cakmak, P. Alves-Oliveira, and J. Gerken. "Multiple Ways of Working with Users to Develop Physically Assistive Robots". 2024. arXiv: 2403.00489 [cs.HC]. URL: https://arxiv.org/abs/2403.00489.

[163] G. Canal, C. Torras, and G. Alenyà. "Are Preferences Useful for Better Assistance? A Physically Assistive Robotics User Study". In *J. Hum.-Robot Interact.* 10, 4 (2021). DOI: https://doi.org/10.1145/3472208. URL: 10.1145/3472208.

[164] F. F. Goldau, M. Pascher, A. Baumeister, P. Tolle, J. Gerken, and U. Frese. "Adaptive Control in Assistive Application - A Study Evaluating Shared Control by Users with Limited Upper Limb Mobility". 2024. arXiv: 2406.06103 [cs.HC]. URL: https://arxiv.org/abs/2406.06103.

[165] M. Selvaggio, M. Cognetti, S. Nikolaidis, S. Ivaldi, and B. Siciliano. "Autonomy in Physical Human-Robot Interaction: A Brief Survey". In *IEEE Robotics and Automation Letters* 6, 4 (2021), 7989–7996. DOI: https://doi.org/10.1109/LRA.2021.3100603.

[166] H. Admoni and B. Scassellati. "Social eye gaze in human-robot interaction: a review". In *Journal of Human-Robot Interaction* 6, 1 (2017), 25–63. DOI: https://doi.org/10.5898/JHRI.6.1.Admoni.

[167] X. Wang. "Eye Gaze-Based Approaches to Recognize Human Intent for Shared Control of Robot Manipulators". PhD thesis. University of California, 2021. URL: https://escholarship.org/uc/item/1wd488hp.

[168] E. H. E. Suryadarma, P. W. Laksono, I. Priadythama, and L. Herdiman. "Controlling Robots Using Gaze Estimation: A Systematic Bibliometric and Research Trend Analysis". In *Journal of Robotics and Control* 5, 3 (2024), 786–803. DOI: https://doi.org/10.18196/jrc.v5i3.21686.

[169] S. W. Brose, D. J. Weber, B. A. Salatin, G. G. Grindle, H. Wang, J. J. Vazquez, and R. A. Cooper. "The Role of Assistive Robotics in the Lives of Persons with Disability". In *American Journal of Physical Medicine & Rehabilitation* 89 (2010), 509–521. DOI: https://doi.org/10.1097/PHM.0b013e3181cf569b.

[170] A. Dünser, M. Lochner, U. Engelke, and D. R. Fernández. "Visual and Manual Control for Human-Robot Teleoperation". In *IEEE Computer Graphics and Applications* 25 (2015), 22–32. DOI: https://doi.org/10.1109/MCG.2015.4.

[171] S. Herr, T. Gross, M. Gradmann, and D. Henrich. "Exploring Interaction Modalities and Task Allocation for Household Robotic Arms". In *Proceedings of the 2016 CHI Conference Extended Abstracts on Human Factors in Computing Systems*. CHI EA '16. San Jose, California, USA: Association for Computing Machinery, 2016, 2844–2850. DOI: https://doi.org/10.1145/2851581.2892456.

[172] J. M. Beer and W. A. R. Akanksha Prakash Tracy L. Mitzner. *Understanding robot acceptance. Technical Report HFA-TR-1103*. Tech. rep. Georgia Institute of Technology, 2011. URL: https://repository.gatech.edu/server/api/core/bitstreams/51f52ee1-3f4f-42d6-a09f-b3b5704be39a/content.

[173] C.-A. Smarr, A. Prakash, J. M. Beer, T. L. Mitzner, C. C. Kemp, and W. A. Rogers. "Older Adults' Preferences for and Acceptance of Robot Assistance for Everyday Living Tasks". In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 56, 1 (2012), 153–157. DOI: https://doi.org/10.1177/1071181312561009.

[174] M. M. Rahman, N. Banik, M. S. H. Sunny, M. I. I. Zarif, D. Bedolla-Martinez, K. Schultz, S. I. Ahamed, and M. H. Rahman. "Wheelchair-mounted robotic arms: a systematic review of technical design and activities of daily living outcomes". In *Disability and Rehabilitation: Assistive Technology* 20, 7 (2025), 2532–2556. DOI: https://doi.org/10.1080/17483107.2025.2547042.

[175] L. Riek and D. Howard. "A Code of Ethics for the Human-Robot Interaction Profession". In *Proceedings of We Robot*. 2014.

[176] C. A. Stanger, C. Anglin, W. S. Harwin, and D. P. Romilly. "Devices for assisting manipulation: a summary of user task priorities". In *IEEE Transactions on Rehabilitation Engineering* 2 (1994), 256–265. DOI: https://doi.org/10.1109/86.340872.

[177] J. M. Beer, A. D. Fisk, and W. A. Rogers. "Toward a framework for levels of robot autonomy in human-robot interaction". In *Journal of Human-Robot Interaction* 3, 2 (2014), 74–99. DOI: https://doi.org/10.5898/JHRI.3.2.Beer.

[178] R. R. Baptista, N. R. Gerszberg, R. V. Godoy, and G. J. G. Lahr. "MIHRaGe: A Mixed-Reality Interface for Human-Robot Interaction via Gaze-Oriented Control". 2025. arXiv: 2505.03929 [cs.RO]. URL: https://arxiv.org/abs/2505.03929.

[179] C. Mazzola, H. Ali, K. Malinovská, and I. Farkaš. "Toward an Interaction-Centered Approach to Robot Trustworthiness". 2025. arXiv: 2508.13976 [cs.RO]. URL: https://arxiv.org/abs/2508.13976.

[180] M. Cakmak, S. S. Srinivasa, M. K. Lee, J. Forlizzi, and S. Kiesler. "Human preferences for robot-human hand-over configurations". In *2011 IEEE/RSJ International Conference on Intelligent Robots and Systems*. San Francisco, CA, USA, 2011, 1986–1993. DOI: https://doi.org/10.1109/IROS.2011.6094735.

[181] A. D. Dragan and S. S. Srinivasa. "A policy-blending formalism for shared control". In *The International Journal of Robotics Research* 32, 7 (2013), 790–805. DOI: https://doi.org/10.1177/0278364913490324.

[182] D. P. Losey, C. G. McDonald, E. Battaglia, and M. K. O'Malley. "A Review of Intent Detection, Arbitration, and Communication Aspects of Shared Control for Physical Human-Robot Interaction". In *Applied Mechanics Reviews* 70, 1 (2018), 1–19. DOI: https://doi.org/10.1115/1.4039145.

[183] P. F. Edemekong, D. L. Bomgaars, S. Sukumaran, and S. Caroline. "Activities of Daily Living". 2025. URL: https://www.ncbi.nlm.nih.gov/books/NBK470404/.

[184] World Health Organization. "Measuring Health and Disability: Manual for WHO Disability Assessment Schedule (WHODAS 2.0)". Geneva, Switzerland, 2010. URL: https://www.who.int/publications/i/item/whodas-2.0.

[185] M. Chi, Y. Liu, Q. Zhang, and C. Zeng. "A demonstration trajectory segmentation approach for wheelchair-mounted assistive robots". In *CAAI Transactions on Intelligence Technology* 10, 3 (2025), 738–754. DOI: https://doi.org/10.1049/cit2.12358.

[186] M. Hayhoe and D. Ballard. "Modeling Task Control of Eye Movements". In *Current Biology* 24, 13 (2014), R622–R628. DOI: https://doi.org/10.1016/j.cub.2014.05.020.

[187] X. Wang, K. Krishnaswamy, M. E. Cabrera, and M. Cakmak. "Robotic Limb Repositioning with Supervised Autonomy". In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*. HRI '20. Cambridge, United Kingdom: Association for Computing Machinery, 2020, 511–513. DOI: https://doi.org/10.1145/3371382.3378387.

[188] A. Jardón, C. A. Monje, and C. Balaguer. "Functional Evaluation of Asibot: A New Approach on Portable Robotic System for Disabled People". In *Applied Bionics and Biomechanics* 9, 1 (2012), 547302. DOI: https://doi.org/10.3233/ABB-2011-0045.

[189] P. H. Chang and H.-S. Park. "Development of a Robotic Arm for Handicapped People: A Task-Oriented Design Approach". In *Autonomous Robots* 15 (2003), 81–92.

[190] S. Arevalo Arboleda, M. Pascher, A. Baumeister, B. Klein, and J. Gerken. "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 '21. Corfu, Greece: Association for Computing Machinery, 2021, 147–155. DOI: https://doi.org/10.1145/3453892.3458044.

[191] L. D. Riek. "Wizard of Oz studies in HRI: a systematic review and new reporting guidelines". In *Journal of Human-Robot Interaction* 1, 1 (2012), 119–136. DOI: https://doi.org/10.5898/JHRI.1.1.Riek.

[192] Tobii Pro. “Tobii Pro Glasses 3 User Manual”. 2024.

[193] Tobii Pro. “Tobii Pro Lab (Version 1.207)”. 2014. URL: https://www.tobii.com/products/software/behavior-research-software/tobii-prolab.

[194] Kinova inc. “KINOVA Gen3 Ultra lightweight robot”. 2022. URL: https://www.kinovarobotics.com/uploads/User-Guide-Gen3-R07.pdf.

[195] J. Wang and E. Olson. “AprilTag 2: Efficient and robust fiducial detection”. In *2016 IEEE/RSJ International Conference on Intelligent Robots and Systems (IROS)*. 2016, 4193–4198. DOI: https://doi.org/10.1109/IROS.2016.7759617.

[196] N. N. P. K. Janaka, S. Zhao, and S. Sapkota. “Can Icons Outperform Text? Understanding the Role of Pictograms in OHMD Notifications”. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*. CHI ’23. Hamburg, Germany: Association for Computing Machinery, 2023. DOI: https://doi.org/10.1145/3544548.3580891.

[197] G. Lancioni, N. Singh, M. O’Reilly, J. Sigafoos, G. Alberti, O. Troccoli, I. Orlando, and C. Ricci. “Enabling people with intellectual and other disabilities to make verbal requests using cardboard chips with mini objects or pictures and a smartphone”. In *Frontiers in Rehabilitation Sciences* 4 (2023), 1–11. DOI: https://doi.org/10.3389/fresc.2023.1257493.

[198] K. Hirose, D. Chugo, S. Yokota, and K. Takase. “Camera-based localization for indoor service robots using pictographs”. In *2011 IEEE/ASME International Conference on Advanced Intelligent Mechatronics (AIM)*. 2011, 924–929. DOI: https://doi.org/10.1109/AIM.2011.6027131.

[199] T. Nishiyama, H. Hoshino, K. Sawada, Y. Tokunaga, H. Shinomiya, M. Yoneda, I. Takeuchi, Y. Ichige, S. Hattori, and A. Takanishi. “Development of user interface for humanoid service robot system”. In *2003 IEEE International Conference on Robotics and Automation (Cat. No.03CH37422)*. Vol. 3. 2003, 2979–2984 vol.3. DOI: https://doi.org/10.1109/ROBOT.2003.1242048.

[200] B. Yang, J. Huang, X. Chen, X. Li, and Y. Hasegawa. “Natural Grasp Intention Recognition Based on Gaze in Human-Robot Interaction”. In *IEEE Journal of Biomedical and Health Informatics* 27, 4 (2023), 2059–2070. DOI: https://doi.org/10.1109/JBHI.2023.3238406.

[201] E. N. Sardinha, M. Munera, N. Zook, D. Western, and V. R. Garate. “Diegetic Graphical User Interfaces and Intuitive Control of Assistive Robots via Eye-gaze”. In *IEEE/RSJ IROS’25*. 2025. DOI: https://doi.org/10.1109/IROS.60139.2025.11247404.

[202] C. Staub, S. Can, B. Jensen, A. Knoll, and S. Kohlbecher. “Human-computer interfaces for interaction with surgical tools in robotic surgery”. In *2012 4th IEEE RAS EMBS International Conference on Biomedical Robotics and Biomechatronics (BioRob)*. 2012, 81–86. DOI: https://doi.org/10.1109/BioRob.2012.6290850.

[203] Z. Wu, B. Ai, T. Silver, and T. Bhattacharjee. “SAVOR: Skill Affordance Learning from Visuo-Haptic Perception for Robot-Assisted Bite Acquisition”. 2025. arXiv: 2506.02353 [cs.RO]. URL: https://arxiv.org/abs/2506.02353.

[204] Roboflow and Supervision. “Computer Vision Model Leaderboard”. 2025. URL: https://leaderboard.roboflow.com/.

[205] Y. Tian, Q. Ye, and D. Doermann. “YOLOv12: Attention-Centric Real-Time Object Detectors”. 2025. arXiv: 2502.12524 [cs.CV]. URL: https://arxiv.org/abs/2502.12524.

[206] B. Dwyer, J. Nelson, and T. Hansen. “Roboflow (Version 1.0) [Software]”. 2024. URL: https://roboflow.com.

[207] K.-M. Ponomarjova, A. Fischer-Janzen, T. M. Wendt, and K. Van Laerhoven. "WLRI-AD: assistive device dataset for daily living automation". In *Personal and Ubiquitous Computing* 30, 1 (2026), 1–10. DOI: https://doi.org/10.1007/s00779-026-01854-2.

[208] I. Corporation. "ROS Wrapper for Intel(R) RealSense(TM) Cameras". 2023. URL: https://github.com/IntelRealSense/realsense-ros.

[209] M. Kudo. "Graphic design of pictograms focusing on the comprehension of people with intellectual disabilities-The next step in standardization: Pictogram design and evaluation methods". In *Visible Language* 56 (2022), 58–85.

[210] G. Canal, G. Alenyà, and C. Torras. "Personalization Framework for Adaptive Robotic Feeding Assistance". In *Social Robotics*. Cham: Springer International Publishing, 2016, 22–31.

[211] R. Feng, Y. Kim, G. Lee, E. K. Gordon, M. Schmittle, S. Kumar, T. Bhattacharjee, and S. S. Srinivasa. "Robot-Assisted Feeding: Generalizing Skewering Strategies across Food Items on a Realistic Plate". 2019. arXiv: 1906.02350 [cs.RO]. URL: https://arxiv.org/abs/1906.02350.

[212] T. Bhattacharjee, G. Lee, H. Song, and S. S. Srinivasa. "Towards Robotic Feeding: Role of Haptics in Fork-Based Food Manipulation". In *IEEE Robotics and Automation Letters* 4, 2 (2019), 1485–1492. DOI: https://doi.org/10.1109/LRA.2019.2894592.

[213] 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*. Cham: Springer International Publishing, 2021, 331–351.

[214] G. Yu, Y. Hu, and J. Dai. "TopoTag: A Robust and Scalable Topological Fiducial Marker System". In *IEEE Transactions on Visualization and Computer Graphics (TVCG)* 27, 9 (2021), 3769–3780.

[215] A. Fischer-Janzen, T. M. Wendt, and K. V. Laerhoven. "Pictogram Fiducial Marker Set". 2025. URL: https://github.com/AnkeLinus/PictogramFidMarkerSet.

[216] 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". In *IEEE Transactions on Systems, Man, and Cybernetics - Part A: Systems and Humans* 42, 1 (2012), 2–14. DOI: https://doi.org/10.1109/TSMCA.2011.2159589.

[217] Z. Fan, C. Lin, and C. Fu. "A Gaze Signal Based Control Method for Supernumerary Robotic Limbs". In *2020 3rd International Conference on Control and Robots (ICCR)*. 2020, 107–111. DOI: https://doi.org/10.1109/ICCR51572.2020.9344272.

[218] R. Zhang, A. Saran, B. Liu, Y. Zhu, S. Guo, S. Niekum, D. Ballard, and M. Hayhoe. "Human Gaze Assisted Artificial Intelligence: A Review". In *Proceedings of the Twenty-Ninth International Joint Conference on Artificial Intelligence, IJCAI-20*. 2020, 4951–4958. DOI: https://doi.org/10.24963/ijcai.2020/689.

[219] M. Kassner, W. Patera, and A. Bulling. "Pupil: An Open Source Platform for Pervasive Eye Tracking and Mobile Gaze-based Interaction". In *Adjunct Proceedings of the 2014 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. UbiComp '14 Adjunct. Seattle, Washington: ACM, 2014, 1151–1160. DOI: https://doi.org/10.1145/2638728.2641695.

[220] P. Alcantarilla, J. Nuevo, and A. Bartoli. "Fast Explicit Diffusion for Accelerated Features in Nonlinear Scale Spaces". In *Proceedings of the British Machine Vision Conference*. BMVA Press, 2013.

[221] P. F. Alcantarilla, A. Bartoli, and A. J. Davison. “KAZE Features”. In *Computer Vision – ECCV 2012*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2012, 214–227.

[222] S. Leutenegger, M. Chli, and R. Y. Siegwart. “BRISK: Binary Robust invariant scalable keypoints”. In *2011 International Conference on Computer Vision*. 2011, 2548–2555. DOI: https://doi.org/10.1109/ICCV.2011.6126542.

[223] E. Rublee, V. Rabaud, K. Konolige, and G. Bradski. “ORB: An efficient alternative to SIFT or SURF”. In *2011 International Conference on Computer Vision*. 2011, 2564–2571. DOI: https://doi.org/10.1109/ICCV.2011.6126544.

[224] D. G. Lowe. “Distinctive Image Features from Scale-Invariant Keypoints”. In *International Journal of Computer Vision* 60 (2004), 91–110. DOI: https://doi.org/10.1023/B:VISI.0000029664.99615.94.

[225] H. Bay, A. Ess, T. Tuytelaars, and L. Van Gool. “Speeded-Up Robust Features (SURF)”. In *Computer Vision and Image Understanding* 110, 3 (2008), 346–359. DOI: https://doi.org/10.1016/j.cviu.2007.09.014.

[226] E. Mair, G. D. Hager, D. Burschka, M. Suppa, and G. Hirzinger. “Adaptive and Generic Corner Detection Based on the Accelerated Segment Test”. In *Computer Vision – ECCV 2010*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2010, 183–196.

[227] J. Matas, O. Chum, M. Urban, and T. Pajdla. “Robust wide-baseline stereo from maximally stable extremal regions”. In *Image and Vision Computing* 22, 10 (2004), 761–767. DOI: https://doi.org/10.1016/j.imavis.2004.02.006.

[228] OpenCV. “OpenCV: Feature Matching. OpenCV Documentation.” URL: https://docs.opencv.org/3.4/dc/dc3/tutorial_py_matcher.html.

[229] S. A. K. Tareen and Z. Saleem. “A comparative analysis of SIFT, SURF, KAZE, AKAZE, ORB, and BRISK”. In *2018 International Conference on Computing, Mathematics and Engineering Technologies (iCoMET)*. 2018, 1–10. DOI: https://doi.org/10.1109/ICOMET.2018.8346440.

[230] F. K. Noble. “Comparison of OpenCV’s feature detectors and feature matchers”. In *2016 23rd International Conference on Mechatronics and Machine Vision in Practice (M2VIP)*. M2VIP 2016. 2016, 1–6. DOI: https://doi.org/10.1109/M2VIP.2016.7827292.

[231] Z. Pusztai and L. Hajder. “Quantitative comparison of feature matchers implemented in OpenCV3”. In *21th Computer Vision Winter Workshop*. 2016, 1–9. URL: https://www.researchgate.net/publication/314286602_Quantitative_Comparison_of_Feature_Matchers_Implemented_in_OpenCV3.

[232] O. Golovnin and D. Rybnikov. “Benchmarking of Feature Detectors and Matchers using OpenCV-Python Wrapper”. In *2021 International Conference on Information Technology and Nanotechnology (ITNT)*. 2021, 1–6. DOI: https://doi.org/10.1109/ITNT52450.2021.9649278.

[233] Pupil Labs. “Pupil Core Terminology”. 2025. URL: https://docs.pupil-labs.com/core/terminology/#offline-fixation-detection.

[234] A. ten Pas, M. Gualtieri, K. Saenko, and R. Platt. “Grasp Pose Detection in Point Clouds”. In *The International Journal of Robotics Research* 36 (2017), 1455–1473.

[235] Z. Erickson, V. Gangaram, A. Kapusta, C. K. Liu, and C. C. Kemp. “Assistive Gym: A Physics Simulation Framework for Assistive Robotics”. In *IEEE International Conference on Robotics and Automation (ICRA)* (2020).

[236] J. Hughes, T. Plumb-Reyes, N. Charles, L. Mahadevan, and D. Rus. "Detangling hair using feedback-driven robotic brushing". In *2021 IEEE 4th International Conference on Soft Robotics (RoboSoft)*. 2021, 487–494.

[237] N. Dennler, E. Shin, M. Matarić, and S. Nikolaidis. "Design and Evaluation of a Hair Combing System Using a General-Purpose Robotic Arm". 2021. arXiv: 2108.01233 [cs.RO]. URL: https://arxiv.org/abs/2108.01233.

[238] I. Asimov. "Runaround". In *I, Robot*. New York, NY: Doubleday, 1950, 40.

[239] Q. M. Marwan, S. C. Chua, and L. C. Kwek. "Comprehensive Review on Reaching and Grasping of Objects in Robotics". In *Robotica* 39 (2021), 1849–1882. DOI: https://doi.org/10.1017/S0263574721000023.

[240] R. Atienza and A. Zelinsky. "Intuitive Human-Robot Interaction Through Active 3D Gaze Tracking". In *Robotics Research. The Eleventh International Symposium*. Berlin, Heidelberg: Springer Berlin Heidelberg, 2005, 172–181.

[241] J. Leaman and H. M. La. "A Comprehensive Review of Smart Wheelchairs: Past, Present, and Future". In *IEEE Transactions on Human-Machine Systems* 47, 4 (2017), 486–499. DOI: https://doi.org/10.1109/THMS.2017.2706727.

[242] C. Antfolk, M. D'Alonzo, B. Rosén, G. Lundborg, F. Sebelius, and C. Cipriani. "Sensory feedback in upper limb prosthetics". In *Expert Review of Medical Devices* 10, 1 (2013), 45–54. DOI: https://doi.org/10.1586/erd.12.68.

[243] C. Moore, S. Corbin, R. Mayr, K. Shockley, P. Silva, and T. Lorenz. "Grasping Embodiment: Haptic Feedback for Artificial Limbs". In *Frontiers in Neurorobotics* 15 (2021). DOI: https://doi.org/10.3389/fnbot.2021.662397.

[244] R. Sigrist, G. Rauter, R. Riener, and P. Wolf. "Augmented visual, auditory, haptic, and multimodal feedback in motor learning: A review". In *Psychonomic Bulletin Review* 20 (2013), 257–263. DOI: https://doi.org/10.3758/s13423-012-0333-8.

[245] E. Borg, J. Rönnberg, and L. Neovius. "Vibratory-coded directional analysis: evaluation of a three-microphone/four-vibrator DSP system". In *The Journal of Rehabilitation Research and Development* 38 (2001), 257–263.

[246] K. Myles and J. Kalb. "Guidelines for Head Tactile Communication". 2010. URL: https://apps.dtic.mil/sti/pdfs/ADA519112.pdf.

[247] K. Myles, M. S. Binseel, and U. A. R. Laboratory. "The Tactile Modality: A Review of Tactile Sensivity and Human Tactile Interfaces. Aberdeen Proving Ground". In *MD: Army Research Laboratory* (2007). URL: https://apps.dtic.mil/sti/citations/ADA468389.

[248] J. Kangas, J. Rantala, D. Akkil, P. Isokoski, P. Majaranta, and R. Raisamo. "Vibrotactile stimulation of the head enables faster gaze gestures". In *International Journal of Human-Computer Studies* 98 (2017), 62–71. DOI: https://doi.org/10.1016/j.ijhcs.2016.10.004.

[249] F. A. A. Kingdom and N. Prins. "Psychophysics: A practical introduction". San Diego, CA, US: Elsevier Academic Press, 2010.

[250] F. Wichmann and H. N.J. "The psychometric function: I. Fitting, sampling, and goodness of fit". In *Perception and Psychophysics* 63 (2001), 1293–1313. DOI: https://doi.org/10.3758/bf03194544.

[251] M. Thibeault, M. Jesteen, and A. Beitman. “Improved Accuracy Test Method for Mobile Eye Tracking in Usability Scenarios”. In *Proceedings of the Human Factors and Ergonomics Society Annual Meeting* 63, 1 (2019), 2226–2230. DOI: https://doi.org/10.1177/1071181319631083.

[252] Tobii Technology. “Accuracy and precision test method for remote eye trackers”. 2011. URL: https://stemedhub.org/resources/3311/download/Tobii_Test_Specifications_Accuracy_and_PrecisionTestMethod_version_2_1_1_.pdf.

[253] Tobii AB. *Tobii Pro Glasses 2 User’s Manual, Version 1.1.3*. 2020. URL: https://www.tobiipro.com/siteassets/tobii-pro/user-manuals/tobii-pro-glasses-2-user-manual.pdf.

[254] M. Stuart, A. B. Turman, J. Shaw, N. Walsh, and V. Nguyen. “Effects of aging on vibration detection thresholds at various body regions”. In *BMC Geriatrics* 3 (2003). DOI: https://doi.org/10.1186/1471-2318-3-1.