DISSERTATION

Enabling Long-term Human-Robot Interaction: On the Need for Behavior Coordination Systems

conducted in partial fulfillment of the requirements for the degree of a

Doktor der technischen Wissenschaften (Dr. techn.)

supervised by

Ao. Univ.-Prof. Dipl.-Ing. Dr. techn. Markus Vincze

and

Assistant Prof.^{in} Mag.^{a} phil. Dr.^{in} phil. Astrid Weiss

submitted at the

Technische Universität Wien
Faculty of Electrical Engineering and Information Technology

by

Dipl.-Ing. Markus Bajones
[redacted]

Vienna, August 2023 Markus Bajones

Acknowledgment

During the course of crafting this thesis, I have had the honor of collaborating with a multitude of inspiring individuals, whose contributions were absolutely essential in bringing this work to fruition.

I want to start by expressing my heartfelt gratitude to my supervisor Prof. Markus Vincze, whose advice, guidance, and trust in my abilities have been nothing short of inspiring. He gave me the freedom to explore and develop my ideas, and I am deeply thankful for his mentorship, which has shaped the very core of my thesis.

A big, personal thank you goes to my mentors and advisors, Astrid Weiss and Michael Zillich. Your unwavering support and the countless conversations we've had over coffee or countless rounds of table soccer have been indispensable. You've been more than mentors; you've been friends and collaborators in every sense.

To my fellow comrades and honorary member in the Hobbit project, especially David Fischinger, Georg Halmetschlager-Funek, Paloma de la Puente, Astrid Weiss, Daniel Wolf, I can't thank you enough for the incredible teamwork and shared enthusiasm. And to the folks in the Squirrel project, particularly Edith Langer, Tim Patten, Jean-Baptiste Weibel, and Michael Zillich, your camaraderie has made this journey all the more enjoyable.

I want to extend my warm thanks to the anonymous reviewers whose feedback challenged me to grow and improve my work. Your critiques were truly valuable. The connections I've made at conferences, workshops, and those unforgettable summer and winter schools have been a lifeline. To everyone I met during these events, your inspirations, insights, and the feeling that we're all in this together have been nothing short of remarkable. I'm also grateful to the general public, whose substantial support, primarily through various EU and national funding initiatives, has made this work possible.

My family, especially my parents Edith and Josef, and brother Jürgen, have been my rock from childhood until the end of this journey. Your unwavering belief in me has been my motivation.

Finally, an especially heartfelt, personal thank you to my partner and friend, Katharina. Your unwavering encouragement and patience have been my driving force. This journey was made more meaningful with you by my side.

Abstract

The deployment of service robots in private homes has become a promising area of research in robotics, bringing significant changes to how household tasks are managed and enhancing residents' quality of life. However, as these robots become more common, a critical challenge arises in achieving effective behavior coordination to ensure smooth and safe interactions with the people in the household. To achieve long-term acceptance and integration into household routines, personalized interaction, adaptability, and continuous learning capabilities are essential.

Previous research has explored two distinct approaches to deploy service robots in private homes. The first approach involves limiting the robot's tasks to short interactions with individuals, where only a limited set of functionalities are accessible. In contrast, the second approach focuses on creating a knowledge-base that stores enough information for the behavior coordination system to make decisions on tasks based on external input. In the first approach, the robot's capabilities are well-suited for short but relatively complex tasks within a well-prepared environment, yet it lacks the ability to explore long-term interactions without additional effort. Conversely, the second approach presents a contrasting obstacle, involving extended setup due to comprehensive architecture construction for tasks' modeling. Nevertheless, once established, this approach enables the robot to facilitate intricate and adaptable behaviors.

Indeed, the challenges discussed have contributed to the current situation in which mobile service robots are not yet widely deployed in the homes of the general public. Instead, their applications are primarily found in well-controlled environments such as museums, hotels, or care facilities, where autonomous service robots have shown success and gained acceptance. One goal is to overcome the complexities of behavior coordination and enhance their capabilities to meet the diverse and dynamic demands of household environments.

To tackle these challenges, we developed an architectural middle-ground for planning, monitoring, and scheduling a set of straightforward tasks. To assess the effectiveness of our approach, we integrated the proposed architecture into five robots and conducted field trials, involving 18 users for a three-week each.

Our work showcases the successful implementation of a behavior coordination system, effectively controlling mobile service robots for extended periods within private homes. Importantly, our approach goes beyond the limitations of isolated and controlled task execution found in previous research.

However, we acknowledge that the system's adaptability to individual users is still limited, despite the extensive efforts of various research groups and private enterprises searching for a comprehensive and universally applicable solution. This challenge remains, and further advancements and innovations are required to enhance the adaptability of service robots to meet the diverse needs and preferences of individual users in real-world settings.

Preface

Throughout my doctoral studies, I have been deeply fascinated by the potential of long-term Human-Robot Interaction (HRI) and its impact on improving the lives of individuals, particularly the elderly, in private homes and care facilities. Investigating how we can achieve successful long-term HRI has been the central focus of my research.

I am immensely grateful for the opportunity to collaborate with a dedicated and talented group of individuals who worked tirelessly alongside me to conduct the field trials with multiple older participants and a mobile service robot. The results of these trials have been enlightening, but they also highlighted the challenges that still lie ahead. Long-term HRI remains a complex and evolving field, and our findings suggest that there is still much work to be done to make it a reality.

During the process of writing my PhD thesis, I was presented with an opportunity to work on service robots for the elderly, both in their homes and in care facilities, within the industry. Eager to contribute further to this domain, I embraced the opportunity, recognizing the potential impact of industry-driven advancements. However, this undertaking extended the timeline for completing my thesis.

Despite the current limitations that hinder the widespread use of service robots by private users, the progress witnessed and actively contributed to during my research journey has given me hope for the future of this field. I remain optimistic about the advancements that will shape the landscape of service robots for individuals living independently.

This thesis represents not only the culmination of my academic pursuit but also a stepping stone in my ongoing commitment to the advancement of long-term HRI and service robotics.

Contents

1 Introduction **1**
1.1 Problem statement . . . 2
1.2 Contributions . . . 3
1.3 Outline . . . 5
1.4 List of publications . . . 5

2 Related Work **9**
2.1 Care Robots . . . 10
2.2 Field trials in Human-Robot Interaction . . . 15
2.3 Willingness to help a robot . . . 20
2.4 Behavior Coordination Systems in HRI . . . 22

3 Behavior Coordination Systems **25**
3.1 Theory . . . 25
3.2 The Hobbit Behavior Coordination Systems . . . 31

4 Field Trials of Hobbit's Behavior Coordination System **37**
4.1 Introduction of Hobbit . . . 38
4.1.1 Robot Platform and Input Capabilities . . . 40
4.1.2 Autonomous provided behaviors . . . 41
4.2 Field Trials . . . 52
4.2.1 Questionnaires/Scales . . . 55
4.2.2 Interview Guidelines . . . 56
4.2.3 Logging Data . . . 56
4.2.4 Pilot Trials . . . 57
4.2.5 Evaluation strategy . . . 59
4.2.6 Trial Procedure . . . 61
4.3 Technical reliability and Usage . . . 63
4.4 Trial results . . . 65
4.4.1 Insights on Usability . . . 65
4.4.2 Insights on User Acceptance . . . 69
4.5 Discussion and Conclusions . . . 74
4.5.1 Lessons Learned . . . 74

4.5.2 Additional Insights . . . 78
4.5.3 Conclusions . . . 79

5 Investigating Adaptation in Behavior Coordination Systems **81**
5.1 Research Questions . . . 81
5.2 Methodology and Scenario . . . 82
5.2.1 Manipulation of Robot Behavior . . . 83
5.2.2 Experiment . . . 85
5.3 Analysis and Results . . . 88
5.4 Conclusions . . . 94

6 Conclusion **95**
6.1 Discussion . . . 95
6.2 Future work . . . 97

List of Tables **99**

List of Figures **101**

List of Abbreviations **103**

Bibliography **104**

Appendices **118**

Appendix A Interview guidelines **119**

Appendix B Timeline **120**

Appendix C Screening Questionnaire **122**

Appendix D Pre-phase interview **125**

Appendix E Midtrial interview **129**

Appendix F End-of-trial interview - Primary User **130**

Appendix G End-of-trial interview - Secondary User **139**

Appendix H Post-phase interview **142**

Chapter 1

Introduction

The use of service robots in private homes has become a promising and transformative area of research in robotics. It is changing how household tasks are managed and improving residents' quality of life.

However, with the increasing prevalence of these robots, there is a significant challenge in ensuring they interact effectively and safely with human occupants. Extending beyond soley executing simple tasks, the goal to ensure the acceptance and seamless integration of these robots into household routines over the long term is important. It requires personalized interactions, adaptability, and continuous learning capabilities.

In order to progress in the domain of service robots in private residences, it is necessary to explore various avenues of research. One important aspect is refining behavior coordination algorithms to handle complex real-world scenarios. Furthermore, the development of behavior models that possess a comprehensive understanding and responsive capability toward human preferences, gestures, and verbal cues is required.

The main focus of the presented work revolves around achieving effective behavior coordination, personalized interaction, adaptability, and continuous learning capabilities in service robots for successful, long-term HRI in private homes.

Furthermore, conducting longitudinal studies involving extended interactions between service robots and household residents provides valuable insights into how these technologies are accepted and adapted in diverse living environments. These studies shed light on the challenges and opportunities in addressing user expectations, privacy concerns, and ethical implications. This, in turn, guides the development of more user-centric and socially responsible service robots.

The presented research was closely aligned with the objectives of the Hobbit project [Bajones et al.; Fischinger et al. [10; 40]]. Its main aim was to develop and deploy a fleet of small service robots designed for home-based personal robotics, with a special focus on supporting older individuals living alone in their homes.

The overarching goal was to promote independent living and improve the well-being of users.

To achieve this, a user-centered approach was adopted, by conducting multiple workshops in Austria and Sweden to gather user requirements [Fischinger et al.; Körtner et al. [40; 64]]. These requirements were refined, considering technical feasibility and the priorities expressed by senior citizens from Austria, Greece, and Sweden, the countries in which the field trials were later performed.

The main functions of the Hobbit robots revolved around fall prevention and fall detection. They also had additional features like picking up objects to clear the floor and the ability to learn and locate specific items in the environment to reduce the risk of falling. Moreover, the Hobbit robots offered fitness instructions, feedback, and brain training games to enhance physical strength, fitness, and cognitive abilities.

In the initial user study across all three countries, the robots semi-autonomously performed tasks in a controlled environment. Based on the study's outcomes and user experiences, a refined list of user requirements and ways the robots should interact with the users was defined. Subsequently, hardware modifications were made to improve performance, safety, and aesthetics before conducting the field trials in Austria, Greece, and Sweden (Figure 1.1). These trials involved 18 users each, lasting three weeks. During the field trials, participants showed high levels of engagement and a willingness to interact with the service robots. The adaptive assistance provided by the robots, tailored to individual user preferences and needs, was particularly well-received. This indicates the potential to foster user independence and reduce the need for external support.

Overall, these findings highlight the significant promise of the Hobbit project in the field of assistive robotics for home-based personal care, while at the same time present challenges we, as the robotics and HRI communities need to address to support autonomous robots in private residences, performing long-term HRI.

1.1 Problem statement

This thesis focuses on addressing the **design and implementation challenges of a behavior coordination system to facilitate long-term Human-Robot Interaction within domestic environments**, both in a general sense and with specific emphasis on:

Research Question 1 Can long-term Human-Robot Interaction be achieved on a mobile service robot using an adaptive behavior coordination system?

Research Question 2 How should a robot behave to perform tasks based on the Human in the loop principle?

Research Question 3 How can the behavior of a robot be adapted to the user to improve long-term HRI?

Figure 1.1: Hobbit robots used by participants during field trials. Male user pointing to an object on the floor, which the robot should pick up (left, Image source: [Bajones et al. [11]]). Female user operating the robot's Graphical User Interface (right, Image source: [Frennert et al. [44]]).

1.2 Contributions

This thesis is dedicated to enhancing the existing landscape of behavior coordination systems utilized in service robots, with a specific emphasis on their role within the realm of long-term Human-Robot Interaction. By delving into the intricacies of behavior coordination, this research seeks to contribute to the evolution of HRI practices, addressing the challenges and complexities associated with prolonged and meaningful interactions between humans and robots in various contexts.

Behavior Coordination System for long-term HRI

Following an extensive review of existing literature and evaluating available behavior coordination systems, the lack of such systems able to support long-term HRI became apparent. Commonly used systems were only used for time-limited task executions and interactions with people, even though the underlying technologies would not prevent that. Our chosen approach expands the use of Hierarchical Finite State Machines (HFSM), monitored and controlled by a heuristic coordinator, which we initially evaluated for multiple basic tasks as reported in [Bajones et al. [5]], [Bajones et al. [6]], [Wolf et al. [122]]. Based on these, we extended the system to incorporate fully autonomous tasks without needed user input [De La Puente et al. [29]], [de la

Puente et al. [30]] as well as complex tasks which include the necessity for command or information input by the user [Vincze et al. [113]], [De La Puente et al. [31]].

Field Trials of Hobbit's Behavior Coordination System

In the realm of HRI research, there has been a noticeable gap in our understanding of how robots and people interact over extended periods. While there have been some studies that ventured into the realm of long-term interactions, they have primarily been confined to the controlled environments of specialized laboratories.

We recognized the need to explore how robots could seamlessly integrate into people's daily lives within the familiar confines of their own homes. We deployed our robots directly into the private residences of participants [Bajones et al. [10]]. The objective was to gain a deeper and more realistic insight into how robots could become an integral part of people's domestic routines. As well as to evaluate the robot's and the behavior coordinator's reliability, user acceptance of the robot, and patterns of usage [Bajones et al. [11]], [Bajones et al. [7]].

The results of the trials indicate that the Hobbit prototypes were well-received by elderly users. They found the robot easy to handle, and effective in meeting their needs [Pripfl et al. [85]]. One of the major contributions of this work involves not only accomplishments but also the identification of barriers that exist in introducing mobile service robots into private homes. One crucial aspect that is often overlooked, is expectation management, which plays a vital role in user acceptance. While participants in experimental settings might tolerate certain shortcomings in a robot, when the robot is intended as a product to assist individuals, the acceptable failure rate is expected to be much lower [Vincze and Bajones [112]], [Vincze et al. [114]].

Investigating Adaptation in Behavior Coordination Systems

Adaptable robot behaviors are crucial for enhancing the effectiveness and acceptance of robots in various real-world contexts. One fundamental reason for this necessity is the dynamic and unpredictable nature of human environments. The need for adaptable robot behaviors is underscored by the desire for personalized and user-centric interactions, as well as understanding how cultural diversity across the global population impacts people's expectations of robot behavior.

Within our field trials, these adaptations encompass variations in communication style, their proximity to individuals, and their navigation patterns. To refine these behaviors, we developed and evaluated a proof-of-concept algorithm that allows the robots to adjust their actions when they encounter unexpected situations or failures [Bajones et al. [8]].

Within a video-based online survey we explored how cultural differences influence people's expectations of robot behavior [Bajones et al. [9]], which allows us to let the robot adjust within help-seeking tasks.

1.3 Outline

This thesis centers on long-term Human-Robot Interaction (HRI) and the utilization of service robots in private residences.

To commence, an exploration of existing literature concerning care robots, field trials, and user responses to malfunctioning robots is undertaken, as discussed in Chapter 2.

Furthermore, the thesis offers a deeper understanding of the original design of a behavior coordination system, explaining its theoretical foundations and architectural structure in Chapter 3.

Subsequently, the Hobbit project and its robot's design are explored, alongside the results of field trials conducted in private residences, as thoroughly discussed in Chapter 4.

Subsequent to that, the findings of a controlled laboratory study investigating specific human behaviors in response to malfunctioning robots are presented, along with a discussion of potential adaptive functionalities based on these outcomes, as elucidated in Chapter 5.

Lastly, the thesis presents a discussion of significant findings, addresses the complexities inherent in conducting substantial research, and outlines potential directions for future exploration. These directions encompass adaptive behavior coordination, the potential impact of cross-cultural influences, the significance of user-centric design, and the robustness enhancement of long-term Human-Robot Interaction systems, all summarized in Chapter 6.

1.4 List of publications

Parts of the work leading to and included in this dissertation has been previously published in the following papers:

- ***Markus Bajones***, *Andreas Huber, Astrid Weiss, and Markus Vincze* **Towards more flexible HRI: How to adapt to the user?**;
  Citeseer

- ***Markus Bajones***, *Daniel Wolf, Johann Prankl, and Markus Vincze* **Where to look first? Behaviour control for fetch-and-carry missions of service robots**;
  arXiv preprint arXiv:1510.01554, 2015

- ***Markus Bajones***, *Astrid Weiss, and Markus Vincze* **Help, anyone? A user study for modeling robotic behavior to mitigate malfunctions with the help of the user**;
  arXiv preprint arXiv:1606.02547, 2016

- ***Markus Bajones*** **Enabling long-term Human-Robot Interaction through adaptive behavior coordination**;
In 2016 11th ACM/IEEE International Conference on Human-Robot Interaction (HRI), pages 597–598, 2016. https://doi.org/10.1109/HRI.2016.7451874

- ***Markus Bajones***, *Astrid Weiss, and Markus Vincze* **Log data analysis of long-term household trials: Lessons learned and pitfalls**;
Workshop on challenges and best practices to study hri in natural interaction settings. In Proceeding of the International Conference on Human-Robot Interaction, 2016

- ***Markus Bajones***, *Astrid Weiss, and Markus Vincze* **Investigating the influence of culture on helping behavior towards service robots**;
In Proceedings of the companion of the 2017 ACM/IEEE international conference on human-robot interaction, pages 75–76, 2017

- ***Markus Bajones***, *David Fischinger, Astrid Weiss, Daniel Wolf, Markus Vincze, Paloma de la Puente, Tobias Körtner, Markus Weninger, Konstantinos Papoutsakis, Damien Michel, Ammar Qammaz, Paschalis Panteleris, Michalis Foukarakis, Ilia Adami, Danai Ioannidi, Asterios Leonidis, Margherita Antona, Antonis Argyros, Peter Mayer, Paul Panek, Håkan Eftring, and Susanne Frennert* **Hobbit: Providing fall detection and prevention for the elderly in the real world**;
Journal of Robotics, 2018. https://doi.org/10.1155/2018/1754657

- ***Markus Bajones***, *David Fischinger, Astrid Weiss, Paloma De La Puente, Daniel Wolf, Markus Vincze, Tobias Körtner, Markus Weninger, Konstantinos Papoutsakis, Damien Michel, et al.* **Results of field trials with a mobile service robot for older adults in 16 private households**;
ACM Transactions on Human-Robot Interaction (THRI), 9(2):1–27, 2019

- *Markus Vincze*, ***Markus Bajones***, *Markus Suchi, Daniel Wolf, Astrid Weiss, David Fischinger, and Paloma da la Puente* **Learning and detecting objects with a mobile robot to assist older adults in their homes**;
In European Conference on Computer Vision, pages 316–330. Springer, 2016

- *Paloma De La Puente*, ***Markus Bajones***, *Peter Einramhof, Daniel Wolf, David Fischinger, and Markus Vincze* **RGB-D sensor setup for multiple tasks of home robots and experimental results**;
In IEEE International Conference on Intelligent Robots and Systems, pages 2587–2594. Institute of Electrical and Electronics Engineers Inc., 2014

- *Paloma de la Puente*, ***Markus Bajones***, *Christian Reuther, Daniel Wolf, David Fischinger, and Markus Vincze* **Robot navigation in domestic**

environments: Experiences using rgb-d sensors in real homes;
Journal of Intelligent & Robotic Systems, pages 1–16, 2018

- *Markus Vincze, **Markus Bajones**, Markus Suchi, Daniel Wolf, Lara Lammer, Astrid Weiss, and David Fischinger* **User experience results of setting free a service robot for older adults at home**;
Service Robots, page 23, 2018

- *Daniel Wolf, **Markus Bajones**, Johann Prankl, and Markus Vincze* **Find my mug: Efficient object search with a mobile robot using semantic segmentation**;
arXiv preprint arXiv:1404.5765, 2014

- *Markus Vincze and **Markus Bajones*** **What a year of trials with a mobile robot in user homes reveals about the actual user needs**;
In Workshop on the Barriers of Social Robotics Take-Up by Society, held at the 26th IEEE International Symposium on Robot and Human Interactive Communication. IEEE, Leicester, 2017

- *Jürgen Pripfl, Tobias Körtner, Daliah Batko-Klein, Denise Hebesberger, Markus Weninger, Christoph Gisinger, Susanne Frennert, Håkan Eftring, Margarita Antona, Ilia Adami, Astrid Weiss, **Markus Bajones**, and Markus Vincze* **Results of a real world trial with a mobile social service robot for older adults**;
In Proceedings of the Eleventh Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts, 2016

- *Paloma De La Puente, David Fischinger, **Markus Bajones**, Daniel Wolf, and Markus Vincze* **Grasping objects from the floor in assistive robotics: Real world implications and lessons learned**;
IEEE Access, 7:123725–123735, 2019

- *Markus Vincze, David Fischinger, **Markus Bajones**, Daniel Wolf, Markus Suchi, Lara Lammer, Astrid Weiss, Juergen Pripfl, Tobias Koertner, and Christoph Gisinger* **What older adults would like a robot to do in their homes-first results from a user study in the homes of users**;
In ISR 2016: 47st International Symposium on Robotics; Proceedings of, pages 1–7. VDE, 2016

Chapter 2

Related Work

In recent years, there has been extensive research in various domains related to HRI. In this chapter, we delve into several areas of research that have received considerable attention, including *social care robots*, *behavior coordination systems in HRI*, *willingness to help a robot*, and *field trials in HRI*.

Social care robots have emerged as a promising solution to support elderly individuals and individuals with disabilities, helping them to perform tasks and maintain their independence. These robots are designed to provide physical assistance, such as helping with activities of daily living, as well as emotional support, such as providing companionship and social interaction.

Behavior coordination systems in HRI have been developed to enable robots to interact with humans in a natural and intuitive way. These systems facilitate the coordination and communication between humans and robots, enabling them to work together effectively and efficiently.

The willingness of individuals to help robots has also been studied extensively in recent years. This research aims to understand the factors that influence individuals' willingness to provide assistance to robots, as well as the impact that the robot's behavior has on their willingness to help.

Finally, field trials in HRI have been conducted to test the feasibility and effectiveness of HRI in real-world settings. These trials enable researchers to evaluate the performance of robots in naturalistic environments and identify areas for improvement.

In the years following the aforementioned studies, numerous research endeavors have been undertaken by various groups to explore different aspects of the field. While the landscape of research has expanded, it is noteworthy that our original results from the conducted studies remain both valid and novel. Interestingly, no other research group has conducted user studies of comparable magnitude and depth in this specific domain. Our findings continue to hold relevance and contribute unique insights, further emphasizing the significance of our work in advancing understanding and knowledge in the field of study. As such, our research stands as a pioneering contribution, offering valuable perspectives that have not been replicated or surpassed by other research groups thus far. It is important to acknowledge that the work conducted by other research groups in the field has been extensively studied and considered within our research. Throughout the subsequent chapters, where relevant, we have diligently incorporated and analyzed the findings, methodologies, and advancements made by these groups. By engaging with and building upon the existing body of knowledge, we have ensured that our research is situated within the broader context of the field, benefiting from the collective insights and progress made by the wider scientific community. This approach enhances the comprehensiveness and robustness of our work, allowing for a more holistic understanding of the subject and enabling us to contribute novel perspectives and advancements within the domain.

2.1 Care Robots

As early as 1996 [Dario et al. [27]] designed and developed mobile robots to assist elderly at home (MOVAID), patients in a hospital (URMAD) or people in a wheelchair (IMMEDIATE). The URMAD goals were to assist humans with searching objects in the environment, picking them up and carrying them to the user. MOVAID on the other hand was supposed to warm up food in a microwave oven, and clean kitchen surfaces as well as removing dirty sheets from a bed within a user's home. Prototypes of the MOVAID robot have been tested and evaluated in a residential site for disabled people, to inform a survey on acceptability and perceived usability of the system. The design of the IMMEDIATE robot is of a wheelchair equipped with a manipulator arm to pickup objects for the user. These robots are shown in Figure 2.1.

Figure 2.1: Prototypes and mock-ups of MOVAID (left), URMAD (center), and IMMEDIATE (right) robotic platforms (Image source: [Dario et al. [27]])

Pearl (Figure 2.2), developed by [Pollack et al. [83]] in the Nursebot project, is a mobile robot developed to fulfill two main tasks for older adults living in a retirement center. First to give reminders about drinking, eating, and taking medicine and second to guide a person to their appointments.

Figure 2.2: Pearl's ability to project a persona relies on essential elements such as the configurable head, including the size and spacing of her eyes, as well as the shape of its lips. The integrated basket of the robot enables easy transportation of objects by placing them inside it. (Image source: `https://www.cmu.edu/cmtoday/issues/dec-2004-issue/feature-stories/human-health/index.html` *) (left); Charles is outfitted with a diverse array of hardware components, including a differential drive system, SICK laser range finder, and sonar and infrared sensors. (middle); The Healthbot robot incorporates a touch screen, camera, speakers, and microphones to facilitate input and output interactions. It utilizes a Hokuyo LIDAR for navigation, allowing it to effectively perceive its environment. Additionally, the robot includes a blood pressure meter, offering the capability to monitor this vital sign as part of its functionality. (Image source: [Stafford [104]]) (right)*

Charles, a robot based on the Peoplebot platform, was designed by [Kuo et al. [66]] to serve users who need vital signs to be monitored on a regular basis. Their main focus was on the inclusion of a multitude of medical equipment to collect a more comprehensive range of vital signs and to explore the ability to perceive and respond to users using multiple modalities. The ability to manipulate the environment was limited due to the fact that the robot only included a 2-DoF gripper, therefore relying on the users to place an object in the gripper before closing it when an object should be carried by the robot (Figure 2.2).

In the Healthbots project [Jayawardena et al. [57]], multiple robots were produced. The hardware itself is based on a mobile platform by Yujin Robot Company (Korea) and extended to accommodate sensors to measure blood pressure, oxygen saturation, and glucose levels as well cameras and microphones (Figure 2.2). Three different trials (with one, two and four robots at the same time) were conducted in a retirement center in New Zealand. During the trials, the robots provided measurements of blood properties with the included sensors, simple entertainment and brain training functionality as well as fall detection with the help of external sensors (connected via a ZigBee network).

Care-O-Bot [Graf et al. [47]] was developed for safe HRI, fetch-and-carry tasks or patrolling in elderly care facilities. The robot is able to grasp objects using its 7-DoF arm (equipped with a 7-DoF dexterous hand) from the floor, the tray on the robot's front and the area behind the robot. Thus enabling it to assist in cleaning up the environment, and serving water, both extensions of the well-known fetch-and-carry tasks. Disadvantages of the Care-O-Bot are, however, the size (making it difficult to navigate in narrow environments, such as private homes) and the high cost. Later generations of the Care-O-Bot improved with regard to these disadvantages though (Figure 2.3).

Figure 2.3: Care-O-Bot designs from first to fourth generation (left to right). Image sources: https://www.care-o-bot.de/en/care-o-bot-3/history/care-o-bot-i.html https://www.care-o-bot.de/en/care-o-bot-3/history/care-o-bot-ii.html https://www.care-o-bot.de/de/care-o-bot-3/download/images.html, https://www.care-o-bot.de/de/care-o-bot-4/download/images.html

The Scitos A5 robot from Metralabs is a mobile robot with autonomous self-charging capabilities. This makes it ideal for long-term use during studies or commercial deployment. It features input through speech recognition or a touch-screen mounted on its chest. On the top of the robot a transparent head is installed, which contains the mechanics for a simplistic face-like structure, with the most prominent features being the actuated eyes (Figure 2.4).

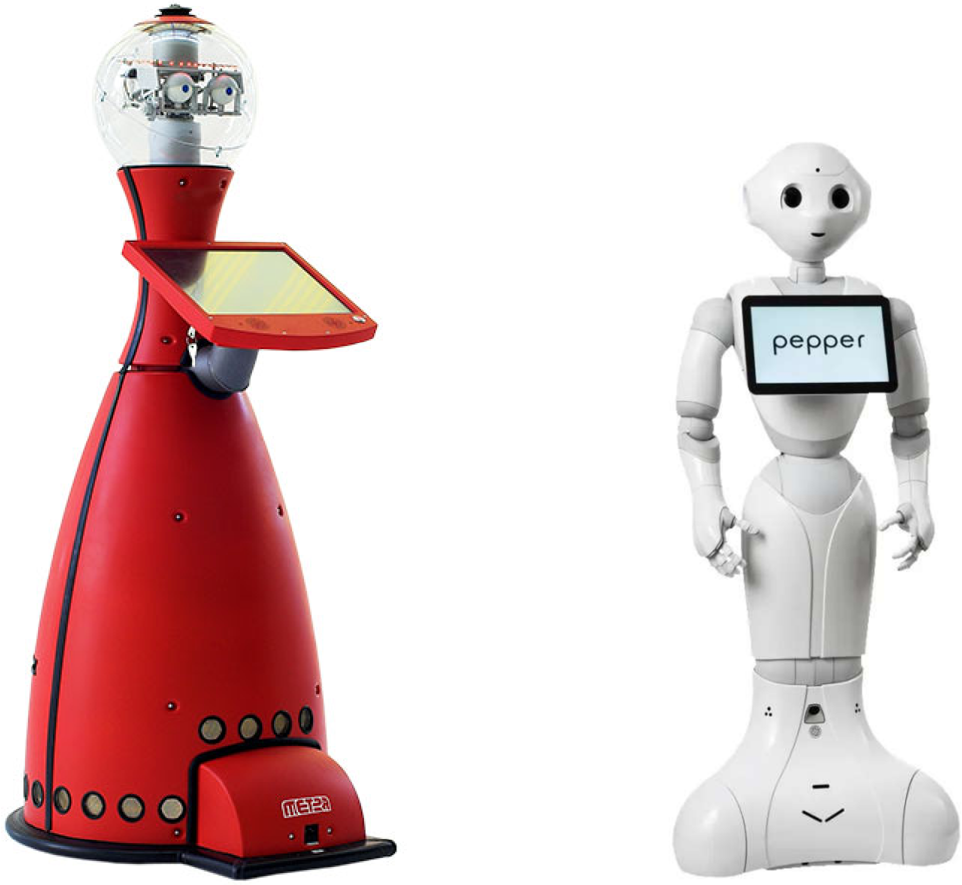


Figure 2.4: The MetraLabs A5 robot is a versatile and mobile robot platform designed for various applications in indoor environments. It is known for its compact size, maneuverability, and advanced navigation capabilities. It is equipped with a customizable payload platform that allows for the integration of different modules or accessories, enabling the robot to perform a wide range of tasks such as delivery, inspection, or data collection. The A5 robot is designed to be user-friendly, providing a flexible and adaptable solution for automation and assistance in diverse settings such as retail, hospitality, healthcare, and research. (Image source: `https://www.metralabs.com/en/service-robot-scitos-a5/`) (left); Pepper finds applications in a range of domains, including retail, hospitality, education, and healthcare. In retail environments, Pepper can assist customers, provide information about products, and even facilitate entertainment or promotional activities. In educational settings, Pepper can serve as a language tutor, engage children in interactive activities, or deliver educational content. (Image source: `https://www.aldebaran.com/en/pepper`) (right)

Pepper is a humanoid robot developed by United Robotics Group, formerly SoftBank Robotics and Aldebaran Robotics (Figure 2.4). It is designed to interact with humans in a natural and engaging way. Standing at about 1.2 meters tall, Pepper is equipped with a variety of sensors, cameras, microphones, and speakers to perceive and respond to its environment.

One of the key features of Pepper is its ability to recognize and interpret human emotions through facial expressions, voice tone, and gestures. This allows it to adapt its behavior and responses accordingly, making interactions more personalized and engaging.

With its humanoid appearance and advanced capabilities, Pepper aims to bridge the gap between humans and robots, offering a socially interactive and helpful companion for various contexts and purposes.

PAL Robotics developed a wide range of mobile robots in the last few years. Most notably are the models ARI and TIAGo (Figure 2.5). ARI, a humanoid robot, is a highly adaptable and versatile robotic platform designed for a wide range of multi-modal expressive gestures and behaviors. It excels as a social robot, suitable for HRI, perception, cognition, navigation, and interaction. Its humanoid form factor adds a human-like touch to its interactions, allowing for enhanced engagement and communication with humans. TIAGo on the other hand is a mobile, modular platform that can be equipped with up to two arms and optional force torque sensors on the wrist. The arms can be used in HRI scenarios as well as in manipulation tasks.

Figure 2.5: PAL Robotics TIAGo (left) and ARI (right). TIAGo is designed for a variety of applications, such as logistics, retail, and education, ARI is useful in education, healthcare, and customer service applications. (Image sources: https://pal-robotics.com/robots/tiago/, https://twitter.com/PALRobotics/status/1463449099773095944)

Kompai is a social robot designed to provide companionship and assistance to the elderly and individuals with disabilities (Figure 2.6). It is specifically developed to address the social isolation and care needs of older adults. The Kompai robot incorporates features such as voice interaction, facial recognition, and telepresence capabilities to engage with users and offer various functionalities like medication reminders, entertainment, and communication with caregivers or family members. It aims to enhance the well-being and quality of life for individuals who may require assistance or companionship in their daily lives.

Figure 2.6: Three Kompai generations (left to right). Image sources: ***https://spectrum.ieee.org/robosoft-kompai-robot-assist-elderly-disabled https://kompai.com/old/ https://www.24presse.com/sante_kompai___robot_for_healthcare_facilities-search-9922646-1-Sante.html/***

Robots with a focus on providing companionship and close physical contact include Paro [Shibata and Tanie [98]], AIBO and Huggable [Stiehl et al. [105]]. These robots are often used to provide mutual companionship with little or no navigation and manipulation skills, thus being distinctively different from the former mention robots that provide at least some kind of functionality to move, and manipulate objects in the environment (Figure 2.7).

2.2 Field trials in Human-Robot Interaction

Multiple projects aimed at developing and deploying assistive, service robots into the homes of users in the last years. [Wada et al. [117]], [Wada et al. [118]], and [Wada and Shibata [116]] focused their work on psychological, physiological and

Figure 2.7: Paro, AIBO (the latest generation as of 2023), and Huggable robots (left to right). Image sources: http://www.parorobots.com/photogallery.asp https://us.aibo.com/ https://aboutdigitalhealth.com/2019/09/17/creepy-cyborgs-or-therapists/

social effects that Paro has on elderly in supported care facilities during long-term studies.

In 2006, [Forlizzi and Disalvo [42]] used Roomba[1] in 14 households. Even though this study used a rather simple robot, offering only little functionality, the results gave insight on how a service robot's introduction into a household changes the view upon such technology and how the introduction of robots offering more functionality needs to consider the social norms within the domestic environment they are placed into.

[Sung et al. [108]] explored the effect of possible customization and therefore passive adaptation to the users in a study within 30 households for six months each. They based their study on the well-known and simple behaving Roomba and found that the possibility to adapt the robot's appearance in such a way that it blends into the environment, can enhance the experience with a robot.

The *Companionable* project [Schroeter et al. [95]] prepared two domestic-like environments, outfitted with Ambient Assisted Living (AAL) technology, and invited six participants (accompanied by five secondary users) to spend two days in one of these homes. Their robot (Scitos G3 platform[2]), with the ability to move autonomously within a well-prepared environment, could only be used freely for up to six hours per day. The study itself focused on participants with mild cognitive impairments, and showed that the functionality to provide reminders and the ability to locate a person was especially well-received by the five primary users.

The *ALIAS* robot [Rehrl et al. [87]], was set up to mainly explore multi-modal input possibilities for HRI, including Automatic Speech Recognition (ASR) and a Brain Computer Interface (BCI). In field trials, lasting two days, with five participants an emergency call, a game, and an e-ticket event booking were tested and the navigational features including human-aware navigation adaptation were demonstrated.

In what can be considered as one of the first long-term user trials with a mobile companion robot in user's homes, the *DOMEO* project [Fazekas et al. [36]] placed

[1] https://www.irobot.com/For-the-Home/Vacuuming/Roomba.aspx
[2] https://www.metralabs.com/en/

a robot in the homes of four elderly users for a total of 287 days. Even though they used a mobile robot, it rarely made use of its ability to move within the home, as indicated by their results.

The *Serroga* project [Gross et al. [49]], direct successor of *Companionable*, moved from the constraints of their AAL equipped living labs into users' homes. They successfully tested their robot, with nine users for a total duration of 16 days in which the elderly users could use the robot for multiple hours without any guidance or intervention.

[Heerink et al. [51]] explored the influence of social abilities, social presence and perceived enjoyment for 30 older adults (aged 65 to 94) using the iCat[3], a stationary robot with the ability to provide facial expressions. They reported that a robot with more social abilities leads to a higher intention to further use it.

The recently finished *MARIO* project [Felzmann et al. [37]] focused on people with dementia living in care facilities. The robot provided a variety of applications on the robot's touch interface and recorded the patient's motion behavior for later analysis. A focus of the project was to assist caregivers in comprehensive geriatric assessment during trials in care facilities in three countries.

[Broadbent et al. [17]] investigated the effect of six stationary robots in multiple aged care facilities. The robots (Guide[4] and Cafero [5]) were deployed for three months in total during which the robots were switched on between 6AM and 8PM. The robots provided entertainment, health-monitoring, and communication functionality to users in the intervention group, while the control group (in different care homes) did not have any contact with the robots. Due to the limited amount of interactions (that needed to be sought-after and started by the users) the amount of collected data did not seem sufficient to show a significant effect between the two groups.

Studies such as the work of [Smarr et al. [102]] explored if older adults would be willing to accept assistance for tasks provided by a robot instead of a human. The investigated tasks were split into self-maintenance Activities of Daily Living (ADL), (e.g. feed, dress, bathe), Instrumental Activities of Daily Living (IADL), (e.g. prepare food, shop, do laundry) and Enhanced Activities of Daily Living (EADL), (e.g. engage in hobbies, learn new skills, communicate for social reasons). After showing the participants videos of a WillowGarage PR2 robot[6], explaining the robot and its abilities in a best-case scenario, the participants showed a preference for a robot assisting in housekeeping tasks (IADL), laundry (IADL), being reminded to take medication (IADL), new learning (EADL), and hobbies (EADL). For ADL tasks, on average, the participants preferred a human assistant to a robot.

[3] `http://www.hitech-projects.com/icat/`
[4] `http://www.ed.co.kr/eng/`
[5] `http://yujinrobot.com/eng/`
[6] `http://www.willowgarage.com/pages/pr2/overview`

In different studies, [Shiomi et al. [99]] and [Kanda et al. [60]] placed multiple robots in a science museum and in a shopping mall. They used different robots[7] for guiding visitors through an exhibition, interaction (playing rock, paper, and scissor, hugging, or handshaking), and bidding farewell. Out of about 10,000 visitors who interacted with the robots, a third returned questionnaires. Their evaluation suggests that the provided guidance increased the interest in science and technology. In the shopping mall a stationary Robovie, using pressure plates to detect visitors and partly controlled by a remote operator for speech recognition and behavior selection / decision-making was used to provide information to visitors. Questionnaires from 235 participating visitors indicated increased usefulness of given information, interest in shops, visiting and shopping frequency when interacting with the robot compared to an information display.

In a more recent study [Broadbent et al. [18]] deployed IRobi[8] in the homes of 30 participants for a duration of four months. The robots' main tasks were the measurement of a set of vital signs and to present Clinical COPD Questionnaire (CCQ)[9], provide reminders, and *I am feeling unwell* function. There findings suggest that such a robot may be useful for Chronic Obstructive Pulmonary Disease (COPD) patients who struggle with keeping their exercise and medication schedule. The authors note however that *A number of technical issues would need to be improved before the robot could be implemented on a larger scale.*

[Sabelli and Kanda [93]] used the human-like Robovie-II platform ([Kanda et al. [59]]) during a one-month period of remote-controlled interaction study in a mall. In this specific study the robot offered random facts about the mall, directions to a specific shop, or a game of *rock, paper, scissors* for children.

In 2018 [Tung and Au [109]] collected data from travel websites for four different hotels which deployed robots for different tasks. The exact models of the robots were not disclosed and only described as humanoid and zoomorphic robots, robotic arm, and robotic butler. Their findings emphasized the impact of robotic embodiment and human-centered perceptions on customer experiences. The findings indicate that users encountered discrepancies when interacting with anthropomorphic robots that closely resembled humans but failed to perform intended human-based tasks, such as assisting with check-in. This led to negative guest experiences and a reluctance to engage with robot services, as they were perceived as mere gimmicks or marketing strategies. Consequently, operators should carefully consider the specific tasks the robot is intended to perform before deciding on its embodiment, whether it be anthropomorphic, zoomorphic, functional, or a combination thereof. Additionally, the study revealed that certain users actively sought out opportunities to engage with robots, aiming to foster a sense of relationship with them.

[Wonsick and Padir [123]] conducted a systematic review of Virtual Reality (VR) interfaces for controlling and interacting with robots. The review included

[7] `https://www.vstone.co.jp/english/products/robovie_x/`
[8] `http://en.yujinrobot.com/archives/portfolio-items/irobi-q`
[9] `http://ccq.nl/`

Table 2.1: Comparison of related user- and field trials (including Hobbit)

	Robotic platform	scheduled vs. Ad hoc interactions	Navigation/ manipulation capabilities	Duration	Environment	Provided functionalities
[Wada et al. [117]] [Wada et al. [118]] [Wada and Shibata [116]]	Paro	Ad hoc	no/no	one year, five weeks	care facilities	companionship
[Forlizzi and Disalvo [42]]	Roomba	Ad hoc	yes/no	three to six weeks	households	floor cleaning
[Shiomi et al. [99]]	Robovie	Ad hoc	yes/no	2 months	science museum	reading RFID tags, physically guiding people to exhibitions
[Heerink et al. [51]]	iCat	scheduled	no/no	10 minutes	lab	weather forecast, jokes, scripted conversations
[Kanda et al. [60]]	Robovie	Ad hoc	no/no	25 days	shopping mall	advertisement, verbally guiding visitors
[Sung et al. [108]]	Roomba	Ad hoc	yes/no	six months	households	floor cleaning
[Fazekas et al. [36]]	Kompai	Ad hoc	yes/no	3 months	households	Emergency call, games, calendar, weather forecast, monitoring of blood pressure and weight
[Rehrl et al. [87]]	Scitos G5	scheduled	yes/no	2 days	-	Emergency call, games, online ticket purchase
[Smarr et al. [102]]	PR2	-	yes/yes	video based study	-	-
[Schroeter et al. [95]]	Scitos G3	scheduled	yes/no	6 hours per day for 2 days	lab	-
[Felzmann et al. [37]]	Kompai	scheduled	yes/no	6 hours per day for 2 days	lab	Smart home integration, cognitive training, calendar
[Gross et al. [49]]	Scitos G3	scheduled	yes/no	8 hours 2-3 hours	lab households	vital sign measurements, reminders, physical activity monitor, calendar
[Broadbent et al. [17]]	Guide, Cafero	Ad hoc	no/no	14 hours per day for 3 months	care facilities	Fall detection, blood pressure, oxygen level measurements, reminders, games
[Broadbent et al. [18]]	iRobi	Ad hoc	no/no	4 months	households	vital sign measurements, reminders
[Zsiga et al. [126]]	Kompai	Ad hoc	yes/no	90 days	households	calendar communication games
[Bodala and Gunes [15]]	Pepper	scheduled	yes/no	5 weeks	lab	mindfulness training
[Li et al. [71]]	Qinmi	scheduled	yes/no	5 months	medical centers	vital signs, radiation dose rate monitoring
[Bajones et al. [11]]	Hobbit	Ad hoc	yes/yes	3 weeks	households	see Subsection 4.1.2

papers published between 2016 and 2020 and categorized them into different areas, including interaction and usability. This indicates that there have been field trials with mobile robots using VR interfaces.

[Gross et al. [48]] presented a comprehensive overview of a long-term field trial with interactive mobile shopping guide robots. The field trials started in 2008 and aimed to study the autonomous operation of robots in everyday environments and their acceptance by customers.

[Lan Hing Ting et al. [68]] conducted fieldwork and field trials in a rehab hospital to co-design a robotic solution for supporting data collection in geriatric assessment. This recent study demonstrates the use of social robots in real-life conditions and highlights the potential for field trials with mobile robots in healthcare settings. Based on these references, it is evident that there have been field trials with mobile robots and users in various domains, including hospitality, shopping, and healthcare. These trials have focused on exploring customer experiences, developing new interaction techniques, and assessing the usability of robotic systems.

In Table 2.1 we show how the Hobbit field trials compare against the state of the art. As can be seen prior work excelled in some shown categories, Hobbit's novelty lies in achieving a combination of successfully testing extensive manipulation capabilities, and the long duration of Ad hoc interactions and autonomous navigation within the user's homes.

The discussed work shows the common balancing act between the duration of the study and the complexity of the robot. When the complexity and the amount of functionalities offered by the robot increased, trade-offs were made, such as (1) reducing the interaction's duration to a few hours (sometimes split over multiple days), (2) carrying out the studies in custom set-up living laboratories, (3) providing a rather small set of functionalities on the robot, (4) reducing the autonomy of the robot, or (5) a combination of some of those trade-offs.

2.3 Willingness to help a robot

A lot of research has already been performed on different techniques for error prevention, such as learning better policies [Argall et al. [3]], shared or sliding control [Shiomi et al. [100]], [Heger et al. [52]], and proactively involving a human to resolve uncertainty in decision-making [Fong et al. [41]], [Nicolescu [79]], and [Rosenthal et al. [90]].

Humans as helpers can have different roles, such as supervisors [Fong et al. [41]], [Shiomi et al. [100]], teachers and demonstrators [Argall et al. [3]], [Nicolescu [79]], and [Hood et al. [54]], and also naive users in the close environment of the robot [Weiss et al. [119]], [Hüttenrauch and Severinson Eklundh [55]]. However, all of this work largely assumes that the robot proactively requests help due to an awareness of limited capabilities (e.g. missing manipulator [Rosenthal et al. [90]]) or missing information (e.g. map knowledge of a specific destination [Weiss et al. [119]]). None of these cases assume that a mistake or a breakdown happened, which needs to be fixed by the user in order to restore the interaction flow.

For instance in the CoBot studies, Rosenthal and colleagues [Rosenthal and Veloso [89]] investigate whether the robot can proactively find people to help with a task, it obviously cannot solve on its own, such as calling the elevator (as it has no manipulator). Asking for help is integrated in the planner with different strategies (identifying if help is needed, how long to wait for help, where to search for help) to enable the robot to reach its goal. In the IURO project Weiss and colleagues [Weiss et al. [121]] studied the acceptance of a robot that proactively asks pedestrians for help to find its way. Asking for the way is therefore in this scenario the intended robot behavior and can be planned in advance. Proactively and reactively asking a user for help, however, have one important aspect in common: The user needs to comply with the robot's request in order to successfully perform the task.

Even though, situations of malfunctioning cannot be planned in advance, suitable recovery strategies can and should be planned in order to keep the interaction with the user alive. An interaction abortion without offering any mitigation strategy negatively impacts the user's perception of the robot [Lee et al. [69]]. It is shown that offering appropriate recovery strategies enables the user to increase the bonding towards failed services [Aaker et al. [1]], [Hart et al. [50]], and [Spreng et al. [103]]. However, it is also proven that people often become emotionally upset when there is a service breakdown, whereby they are more dissatisfied by the failure of the recovery than by the mistake itself [Bitner et al. [14]]. So far little research has been done on mitigation strategies for service robots. Prior research showed that people feel a loss of control when they do not understand why a robot fails [Hinds et al. [53]]. The more autonomous a robot was, the more people blamed it for failure and explaining the reason for failure led only to little improvement [Kim and Hinds [61]]. Moreover, it was shown that people's orientation toward services influences which recovery strategy works best for them. Those with a relational orientation responded best to an apology; those with a utilitarian orientation responded best to a compensation [Lee et al. [69]].

[Vanzo et al. [111]] conducted a user study to assess participants' attitudes towards collaborating with robots. The study involved participants interacting with a robot in a collaborative task. The participants' attitudes and behaviors towards the robot were observed and analyzed. Their findings suggest that people generally have positive attitudes towards collaborating with robots. Participants were willing to interact with the robot and help it in completing the task. However, the study also identified factors that influenced individuals' attitudes and behaviors towards robots. These factors include the domain of application and design of the robot, the type of exposure to the robot, and the characteristics of potential users.

In an online video study conducted by [Daly et al. [26]], the researchers sought to examine observable patterns that could enhance people's inclination to assist a robot in a state of distress. The study specifically focused on investigating the influence of emotional behavior and its ethical implications. The findings revealed that participants demonstrated a greater willingness to offer assistance when the robot displayed atypical behaviors, such as visualizing a happy face.

[Fallatah et al. [35]] propose a model with microcultures as a factor and analyze its effects on people's willingness to help robots. Through a laboratory experiment, participants representing various microcultures were recruited, and bystanders in different coffee shops were tasked with assisting a robot in purchasing coffee. The findings underscore the existence of divergent levels of willingness to help robots, with certain microcultures exhibiting greater levels of helpfulness compared to others. Moreover, the study highlights cultural disparities in participants' perceptions of robots and their levels of engagement with them.

[Naneva et al. [78]] conducted a systematic literature review focusing on the trust levels associated with social robots. The study revealed interesting findings. Although the overall weighted mean for trust was approximately zero, suggesting a general lack of strong trust or distrust towards social robots, the analysis of variations indicated that 43% of the studies provided evidence of people harboring distrust towards these robots. The observed lack of trust towards social robots is likely to hinder the willingness of those individuals to offer assistance or aid to robots in need.

2.4 Behavior Coordination Systems in HRI

Moving towards autonomous service robots, behavior coordination systems constitute an important building block to fulfill the requirements of action planning, safe task execution, and integration of HRI. HAMMER from Demiris and Khadhouri [Demiris and Khadhouri [32]] is built upon the concept to use multiple forward or backward control-loops, which can be used to predict the outcome of an action and compare this against the actual result of the action. Through this design it is possible to choose the action with the highest probability of reaching the desired outcome which has successfully been used in a collaboratively controlled wheelchair system [Carlson and Demiris [22]], in order to correct the user's input to avoid an erroneous situation.

Cashmore et al. [Cashmore et al. [23]] introduced ROSPlan, a framework that uses a temporal planning strategy for planning and dispatching robotic actions. Depending on the needs, a cost function can be optimized to plan in a certain manner (e.g. time- or energy-optimized). However, the constructed plan is up until now only available as a sequence of executed actions and observed events, but no direct focus is put on the human, besides modeling the user as means to acquire some event (e.g. moving an object from one location to another).

Mansouri et al. [Mansouri and Pecora [72]] incorporate temporal and spatial reasoning in a robot tasked with pick and place in environments suited for users. In the context of ALIAS, Goetze et al. [Goetze et al. [46]] designed their *dialogue manager* for the tasks of emergency call, a game, e-ticket event booking, and the navigation as state-machines. However, there are still significant research challenges regarding how to incorporate humans into the planning stages and decide when the robot needs to adapt to the user instead of staying with the planned task.

Most of these systems either treat the human as an essential part of the system [Carlson and Demiris [22]] (e.g. for command input) and rely on the user to execute actions planned by the coordination system [Koppula et al. [62]]. Such systems only work under the precondition that the robot will execute a given task for the user independently of the user input [Mansouri and Pecora [72]].

In their comprehensive investigation, [Ghzouli et al. [45]] undertook a thorough survey of behavior models available on Github. Their search criteria (compatibility with ROS and the distinction between models used primarily for instructional purposes) narrowed the behavior models down to finite state machines and behavior trees. State machines offer a structured approach to modeling behavior, where the robot's actions are defined based on discrete states and transitions between them. On the other hand, behavior trees provide a hierarchical representation of behaviors, allowing for flexible and modular design.

In their work, [Cao et al. [20]] introduces a collaborative homeostatic-based behavior controller that encompasses two distinct layers: a reactive heuristic layer and a deliberative layer. The controller operates by leveraging sensory input from these layers to generate, execute, and monitor a plan aimed at accomplishing a specific goal.

The reactive heuristic layer serves as the initial processing stage, where predefined rules and reactive behaviors are employed to handle immediate sensory input. This layer enables rapid responses to environmental stimuli, allowing the system to react and adapt to changing circumstances.

On the other hand, the deliberative layer takes a more thoughtful and strategic approach. It generates a plan based on the given goal and evaluates various possibilities to determine the most suitable course of action. This layer considers both the current state of the system and the desired outcome, taking into account the information provided by the reactive and reflective layers.

The reflective layer plays a crucial role in the behavior control architecture. It acts as an evaluator, continually monitoring and assessing the plan generated by the deliberative layer. If deemed necessary, the reflective layer can intervene and modify the plan to optimize the chances of successfully achieving the desired goal. This capability enables the system to dynamically adapt its behavior based on real-time feedback and environmental conditions.

By incorporating both reactive and deliberative layers, Cao's behavior controller achieves a balance between instantaneous reactions and thoughtful decision-making. This collaborative approach allows for flexible and efficient behavior control, ensuring that the system can respond promptly to immediate stimuli while still maintaining a coherent and goal-oriented behavior plan.

In a subsequent work, [Cao et al. [21]] introduced a behavior control system tailored for social robots used in therapy settings, emphasizing personalization and platform independence. The primary objective of this system is to enhance autonomy and alleviate the human workload involved in healthcare interventions.

The system architecture enables the robot to exhibit a personable character, where its behaviors are customized based on user profiles and responses during HRI. These behaviors are designed at abstract levels, allowing for seamless transferability across various social robot platforms.

To realize this architecture, their work adopts a component-based software engineering approach. This approach ensures that the developed components can be easily replaced and reused, promoting flexibility and modularity within the system. By employing this methodology, the behavior control system achieves greater adaptability and extensibility, facilitating the integration of new features or the incorporation of different social robot platforms.

Through the integration of personalization, platform independence, and a component-based software engineering approach, their behavior control system advances the field of social robotics in therapy. By reducing the human workload and promoting autonomy, the system paves the way for more effective and engaging healthcare interventions that leverage the capabilities of social robots.

Chapter 3

Behavior Coordination Systems

A behavior coordination system is an essential component of any robot that interacts with humans. The primary purpose of a behavior coordination system is to enable the robot to exhibit appropriate behavior in response to various environmental stimuli and human actions. This system allows the robot to coordinate its actions and behaviors with those of humans, making its behavior more predictable and understandable to humans.

One of the key advantages of a behavior coordination system is that it can enable robots to adapt to different situations and interact with humans in a more natural and intuitive manner. By analyzing human behavior and responding appropriately, robots can become more effective at performing tasks and providing assistance to humans.

Another important benefit of behavior coordination systems is that they can improve the safety and reliability of robots. By coordinating their actions with humans, robots can avoid collisions and other potential hazards, reducing the risk of accidents and injuries.

Overall, the need for a behavior coordination system in a robot arises from the fact that robots are designed to interact with humans in various settings, ranging from domestic to industrial environments. To ensure that these interactions are safe, effective, and beneficial to humans, robots need to exhibit appropriate behavior and respond appropriately to various stimuli and actions. A behavior coordination system is the key to achieving these goals, enabling robots to interact with humans in a natural and intuitive way, while also ensuring their safety and reliability.

3.1 Theory

Behavior Coordination Systems are typically classified as one of *hierarchical*, *reactive or behavior-based*, and *hybrid* paradigm. For better understanding the biological concept of *behaviors* needs to be discussed.

Biological concept of Behaviors

A *behavior* represents a mapping between sensory input to a pattern of motor actions which then are used to achieve a task [Brooks [19]].

Figure 3.1: Behavior is a mapping between sensor input and motor action of actuators.

Reflexive The output of this behavior is hard-wired between the sensor input and the according response. Reflexive behaviors have been classified into *Reflexes*, *Taxes*, and *Fixed-action patterns*. [Arkin et al. [4]], [Beer et al. [13]], and [McFarland [74]] defined them as

Reflexes are rapid, automatic involuntary responses triggered by a certain environmental stimulus. The reflexive response persists only as long as the duration of the stimulus. Further, the response intensity correlates with the stimulus' strength. Reflexes are used for locomotion and other highly coordinated activities.

Taxes are behavioral responses that orient the animal toward or away from a stimulus. Taxes occur in response to visual, chemical, mechanical, and electromagnetic phenomena in a wide range of animals.

Fixed-action patterns are time-extended response patterns triggered by a stimulus but persisting for longer than the stimulus itself. The intensity and duration of the response are not governed by the strength and duration of the stimulus, unlike a reflex. Fixed-action patterns may be motivated, unlike reflexes, and they may result from a much broader range of stimuli than those that govern a simple reflex. Motivated behaviors are governed not only by environmental stimuli but also by the internal state.

Reactive These behaviors are learned over time and can be executed without the need to plan their execution. Reactive behaviors are trained by repeating the action over a long period of time until they are inscribed as so-called *muscle memory* within the brain.

Conscious Are well-designed combinations of already developed behaviors. This means that while no planning is involved a desired action can be achieved through combinations of behaviors from any of the previously discussed categories.

Hierarchical paradigm

Is the oldest method dating back to the robot Shakey [Nilsson [80]], on which development started in 1966. The *hierarchical paradigm* consists of the stages *Sense*, *Plan*, and *Act*. Figure 3.2 shows this structure.

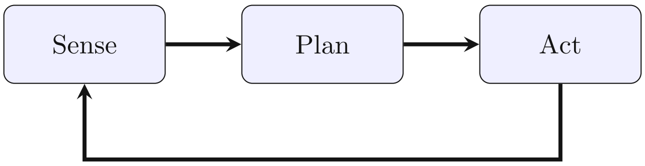


Figure 3.2: Hierarchical paradigm

During the *Sense* stage the robot's sensors collect data of the surrounding environment. In *Plan* the collected data and the currently known state is used to update the current model of the world and plan the next step of the robot to achieve its current goal. The *Act* stage is when the actual execution of the planned action happens. Afterward this cycle is repeated until (a) the goal has been achieved or (b) the goal has been modified. Problems with this paradigm are the computational burden while planning and the difficulties of maintaining a world model.

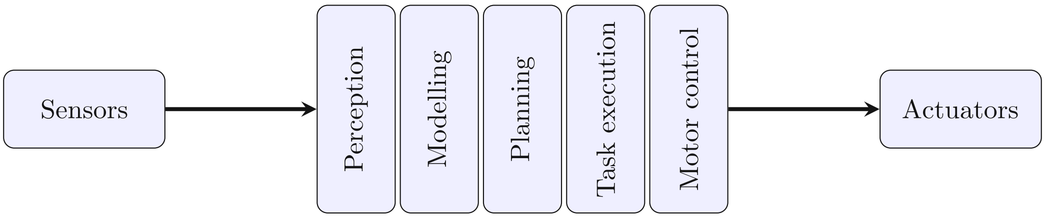


Figure 3.3: Horizontal decomposition according to [Brooks [19]]

Figure 3.3 visualizes the horizontal decomposition of the hierarchical paradigm.

Reactive or Behavior-based paradigm

This paradigm is based on *behaviors*, biological inspired connections between *Sense* and *Act* stages ([Arkin et al. [4]]). The sensor input is directly linked to an action, which allows for fast execution (Figure 3.4) and form the building blocks of complex actions for the robot. As there is no explicit planning stage in the *reactive* paradigm all actions emerge from the execution of multiple parallel behaviors. The responsibility to select which of the individual behaviors are active at a specific time falls into the design considerations of the architecture.

Compared to the hierarchical paradigm and its horizontal decomposition, the behavior-based shows a vertical decomposition. This means that all sensing is happening on a local (at the behavior) basis, and the result of each can contribute to the full outcome.

Figure 3.4: Behavior consists of a pair of Sense *and* Act *stages*

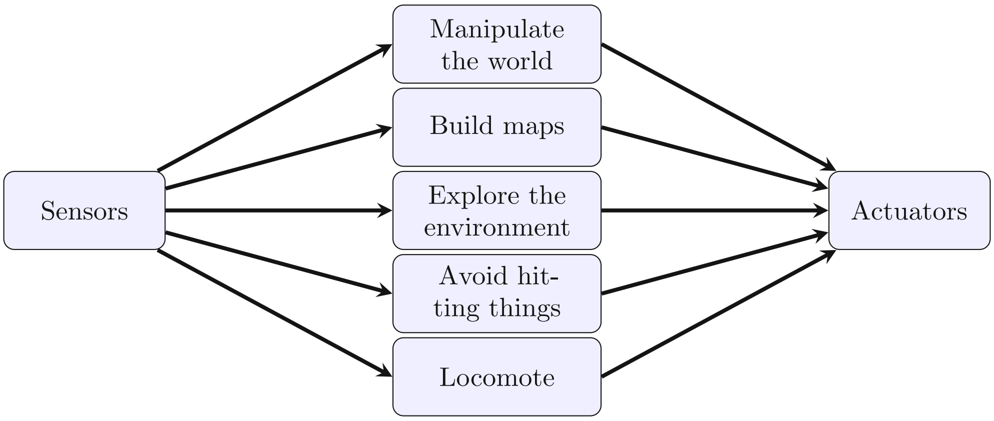


Figure 3.5: Vertical decomposition according to behavior-based paradigm

The *Subsumption* is one of the prominent underlying methods that define how individual reactive behaviors are combined into a robot's overall behavior. In Figure 3.6 the different layers of competence are represented. On each layer the behavior is executed without knowledge of an internal state or of the execution of other layers. However, the outcome from a higher layer can alter the actions of a lower by either *Suppression* or *Inhibition*. *Suppression* is the mechanism in which the input signal to a behavior is replaced by the output signal of the subsuming layer. *Inhibition* on the other hand connects the output signals of a behavior and the output of the subsuming layer in such a way that the behavior output can be turned on or off. While the behavior-based architecture has advantages such as the modularity and its ability to test behaviors independently, and the fact that it is not necessary to maintain a complete world model, it also falls short in its ability to plan and reason about the environment.

Another method is *Motor schemas* that represent the behavioral responses in form of vectors generated by a potential fields approach. The building blocks of behaviors are *Motor schema*, or memories of movement parameters and *Perceptual schema*, or perceived consequences of movements [Wulf [124]] (Figure 3.7). The author of the *schema* theory argue that a generalized motor program with a set of motion commands is retrieved from memory and adapted for the given situation [Schmidt [94]].

The primitives to calculate a full behavior are *Uniform*, *Perpendicular*, *Attraction*, *Repulsion*, and *Tangential*. These primitives are summed up in the form of vector addition for the overall behavior.

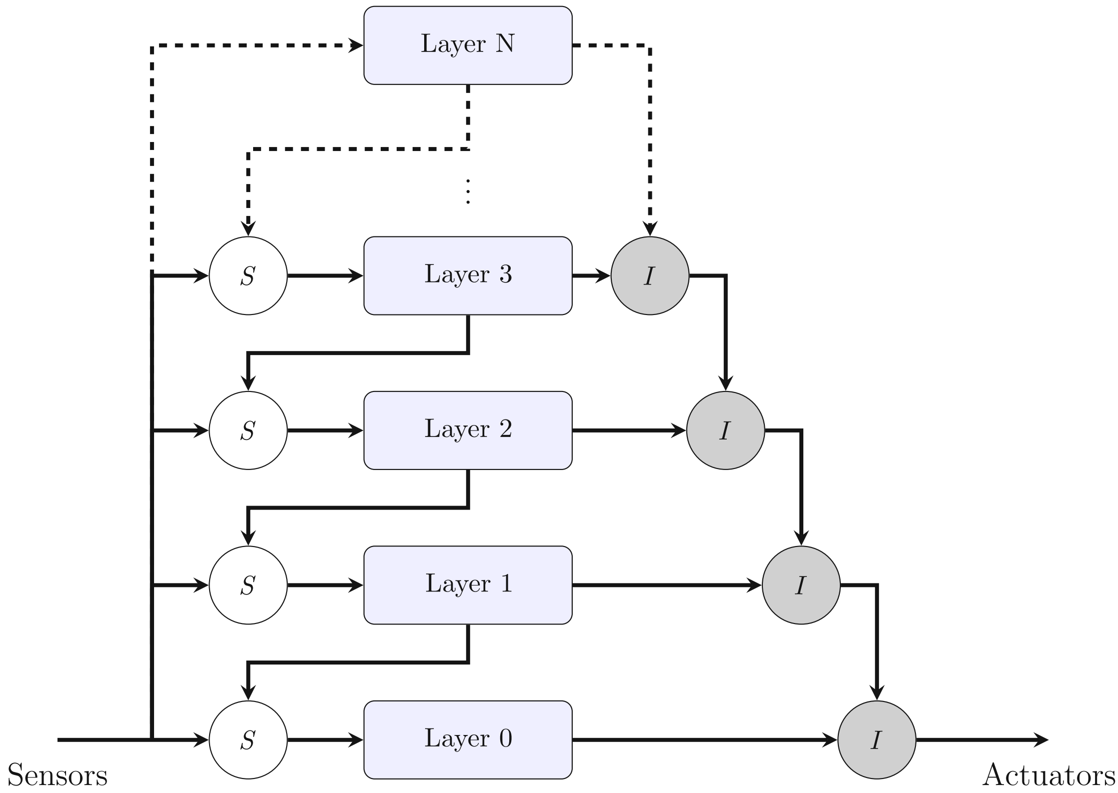


Figure 3.6: Behavior-based paradigm consists of concurrent combinations of behaviors out of which the global behavior emerges. S *denotes suppression of an input signal to a lower layers behavior.* I *or inhibition is the mechanism of overwriting a lower layers output signal. Adapted from [Briot et al. [16]]*

Hybrid paradigm

The hybrid paradigm combines the hierarchical and the reactive architectures to take advantage of long-term planning as well as reactive behaviors on the robot. Instead of the traditional iteration over the three stages, the execution starts with the *Plan* stage, and perform *Sense* and *Act* stages afterward at the same time (Figure 3.8).

The idea to decouple planning stems from two different assumptions. First, planning algorithms are computationally expensive and should be decoupled to prevent a slow-down of reactive (near realtime) program execution. Second, planning and reactive algorithms are based on different times scales and knowledge scopes (local vs. global), should be encapsulated according to the software development principle of coherence. It is important however to understand that sensing, which happens in parallel to acting, is not limited to behaviors. Thus, sensing feeds the planning stage as well as the local behaviors as presented in Figure 3.9.

The main difference to the reactive paradigm is that behaviors within the hybrid approach include innate, reflexive, and learned behaviors. As implementations of

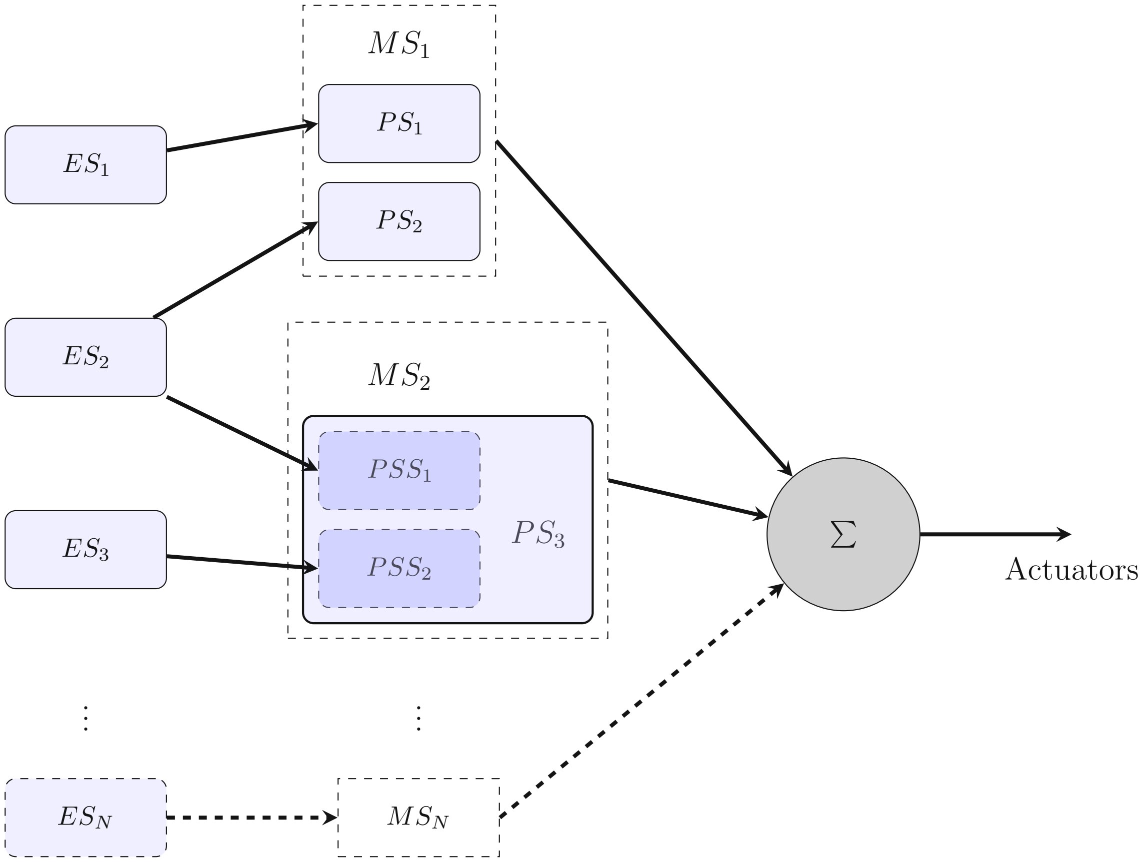


Figure 3.7: Relation between the sensed environment, perceptual and motor scheme, and the robot's actuators. The vector fields representing the individual outputs are added up in the vector sum block to build the overall behavior as is sent to the robot's actuators. ES *represents* Environmental sensors, PS *a* Perceptual schema, PSS *a* Perceptual sub-schema, *and* MS *a* Motor schema. *Adapted from [Ravangard [86]].*

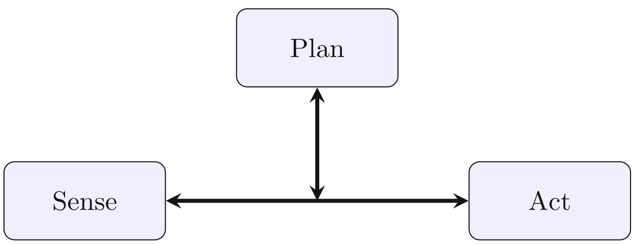


Figure 3.8: Hybrid paradigm

hybrid architectures tend to build more complex and larger combinations of behaviors they are sometimes referred to as skills so that they can easily be distinguished from reflexive behaviors.

To actually make use of these skills *behavioral management* and *performance monitoring* need to be taken into account. *Behavioral management* is the process

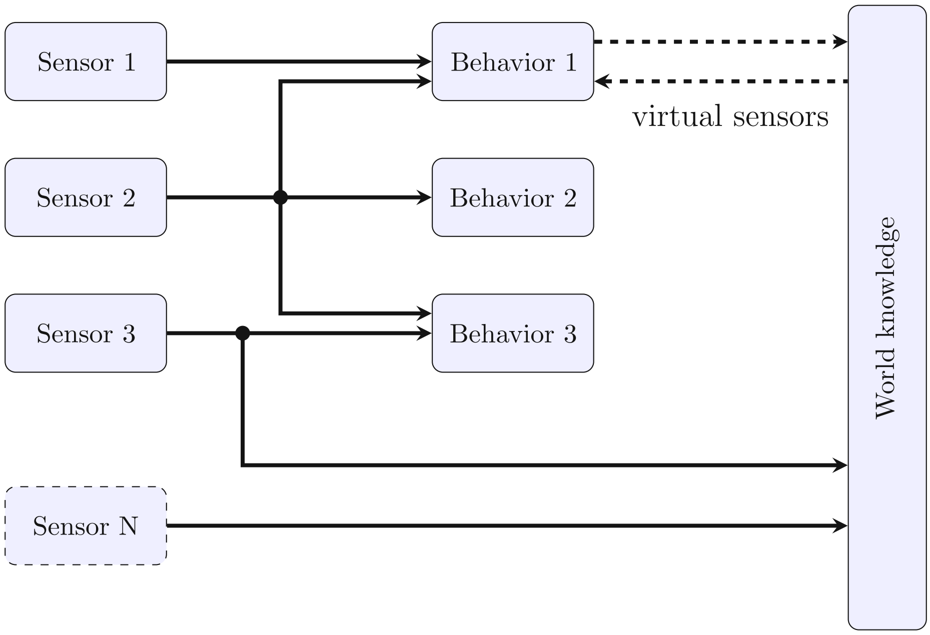


Figure 3.9: Sensing in hybrid architectures

of selecting which set of behaviors to execute in what order. To perform this decision-making process knowledge from outside (global) needs to be collected and analyzed. This knowledge can be collected by i.e. environmental sensors, that collect data such as color or distance information; or from virtual sensors, that senses from within other behaviors (i.e. the calculated current position in the environment). *Performance monitoring* is needed to evaluate if any executed skill actually performs its task and moves towards the given goal. Global knowledge, i.e. the goal and current state information of all active skills need to be accessible to monitor the robot's performance at any given time.

3.2 The Hobbit Behavior Coordination Systems

The architecture, as shown in Figure 3.10, is designed to provide a flexible and adaptable approach to robot task planning and execution. We have employed a hybrid paradigm that combines reactive behaviors with high-level skills and planning to create an efficient and effective system.

The *planning* stage of our system is responsible for long-term task planning and the maintenance of a global world model. This stage uses heuristic planning to determine the robot's overall objectives and then directs these objectives down into skills and tasks that can be executed as reactive behaviors.

The *Sense and Act* stages of our system are responsible for realtime task execution. These stages use reactive behaviors to monitor the environment and

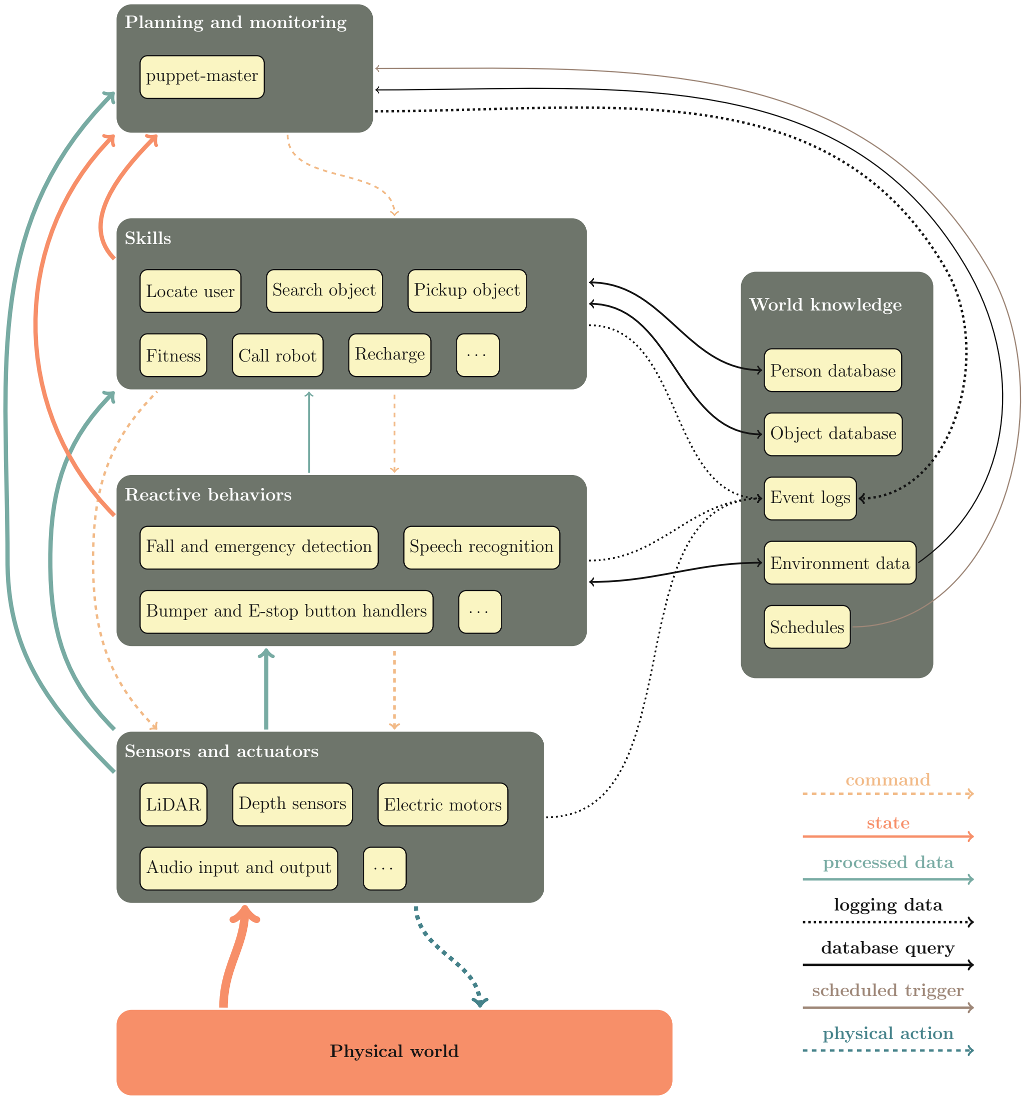


Figure 3.10: Hobbit's behavior coordination system: ***Sensor and Actuator*** *stage interacts directly with the* ***Physical world****, while sending the processed data to and receiving commands from other stages.* ***Reactive Behaviors*** *further processes the incoming data and triggers reactions or publishes the calculated results to higher stages.* ***Skills*** *either receive data from* ***Sensors*** *or pre-processed data from* ***Reactive behavior*** *stages.* ***Planning and monitoring*** *receives the current states of the lower stages, retrieves data from the environment database, and receives scheduled triggers. It further sends commands, i.e. to switch from a running skill to another one, to the* ***Skills*** *stage. All stages send store their logs locally while also sending them to the centralized* ***Event logs*** *database within* ***World knowledge****. The thickness of the arrows indicate the amount of data transfered over a specific communication channel.*

respond to stimuli in as fast as possible. By using this approach, our system is able to achieve both flexibility and reliability in task execution.

In addition to developing the architecture, we prioritized creating independent components that are not tied to a specific robot. This design choice allows for greater flexibility and adaptability, as it enables the integration of additional skills and behaviors as drop-in replacements. This approach also ensures that the existing interface definitions are maintained and utilized by any new components added to the system. By implementing this design principle, our behavior coordination system can be easily tailored to suit different robot platforms, while maintaining its functionality and effectiveness in a variety of scenarios. Overall, our focus on modularity and compatibility enhances the versatility and longevity of the system, ensuring that it can evolve and improve over time.

In a more detailed description, our behavior coordination system is predominantly implemented as a collection of hierarchical finite state machines (HFSMs) across various levels. However, it's important to note that certain levels, specifically those involving hardware-facing sensor input or actuator output code, are not implemented as finite state machines. Instead, they are realized as pass-through nodes that facilitate the transfer of data between the actual signal producer and signal processor. The determination of which part serves which role depends on the direction of data flow between the hardware-related nodes and any node within the HFSM.

Planning and Monitoring Stage

This stage is responsible to determine which skill should be active at any given time. This decision is made by considering various factors such as the priority of the currently active skill, processed input data (e.g. speech commands from a user or emergency detection), status data from the active skill, or an execution request for a scheduled skill. In this stage, a pivotal component known as the "puppet-master" node takes charge of controlling and monitoring the skills within the system. The puppet-master node implements this control, thus it plays a crucial role in orchestrating the activation and deactivation of skills, ensuring the system operates in a coordinated and efficient manner. By leveraging input data and skill priorities, the behavior coordination system can make informed decisions during the planning stage, effectively managing the activation and transition of skills based on the given context and requirements.

Skills Stage

The implementation of skills in the behavior coordination system in the form of a collection of HFSMs was carried out using Python in conjunction with the SMACH library, which provides a comprehensive set of ROS-compatible states. Leveraging preexisting states offered by SMACH greatly facilitated the integration of the architecture into any ROS-enabled robot. Notable examples of these states include the *MonitorState*, used for processing data received on specific ROS topics based on given conditions; the *SimpleActionState*, well-suited for navigation or manipulation tasks; and the *CBState*, providing a simplified way to encapsulate a callback function with minimal code overhead. In cases where the provided SMACH states did not fully meet hardware requirements or align with the existing code and its API, customization was performed by extending or rewriting them to ensure optimal compatibility and performance.

The puppet-master node, operating within the planning stage of the architecture, instantiates the set of skills. In this stage, the execution of any skill can only be initiated, ensuring a controlled activation process. However, interrupting an active skill is possible from both the planning and reactive behavior stages, allowing for dynamic adjustments based on the system's requirements and external stimuli.

Reactive Behaviors Stage

The reactive behaviors stage is essential for near realtime responsiveness and adaptability. Reactive behaviors are implemented at a lower level in the architecture and will inform the active skill in parallel to the reaction's execution. Reactive behaviors ensure that the robot can handle emergency situations, avoid obstacles, or adjust its actions based on immediate sensory input. Examples of such implementation are the handling of wireless buttons used for calling the robot or for emergency notification, detection of a person who has fallen or is falling at the moment, or avoiding a suddenly appearing obstacle while the robot is moving in the environment. The data input for the reactive behavior is either passed directly or preprocessed from the sensor and actuator stage.

Sensors and Actuators Stage

The sensor and actuator stage establishes communication between the robot's physical sensors, actuators, and the active skill. It handles the flow of information between the robot's sensory inputs (such as cameras, microphones, or touch sensors) and the skill that requires access to specific sensory data. Likewise, it enables the skill to control the robot's actuators (such as motors, grippers, or speakers). This stage ensures seamless integration of sensor data and actuation commands with the active skill. Some sensor data is preprocessed within this stage as the raw data is not needed in the high-level skills. Audio input for example is first filtered, processed by speech detection and recognition before the recognized speech (if any) is sent to the other stages.

World Knowledge

World knowledge serves as a repository for storing relevant information about the robot's environment, objects, and experiences. This knowledge can be accessed during the planning and execution stages to enhance decision-making and improve task performance. The architecture uses the world knowledge to make informed decisions, retrieve contextual information, or adapt the active behavior based on previous interactions. Information stored within is used by the robot to look up the location a person has been seen at in the past and how often this was the case. This knowledge is used as a seed when the robot needs to locate a specific person in the environment by planning a navigation path through all locations they were detected by the robot in the past.

Utilizing these stages in the behavior coordination system, the robot gains the ability to plan, execute tasks using specific skills, react in near realtime to dynamic situations, interact with its environment through sensors and actuators, and leverage stored knowledge. This comprehensive approach enables the robot to exhibit complex behaviors, adapt to changing circumstances, and effectively interact with humans and the surrounding world.

It is important to note that the complexity of these Hierarchical Finite State Machines increases proportionally with the number of nested states or Finite State Machines they contain. While modular components can be reused, comprehending the overall system becomes more challenging. An example of this complexity is illustrated in Figure 3.11, which represents an HFSM utilized by the Hobbit robot for the task of learning a new object with user assistance.

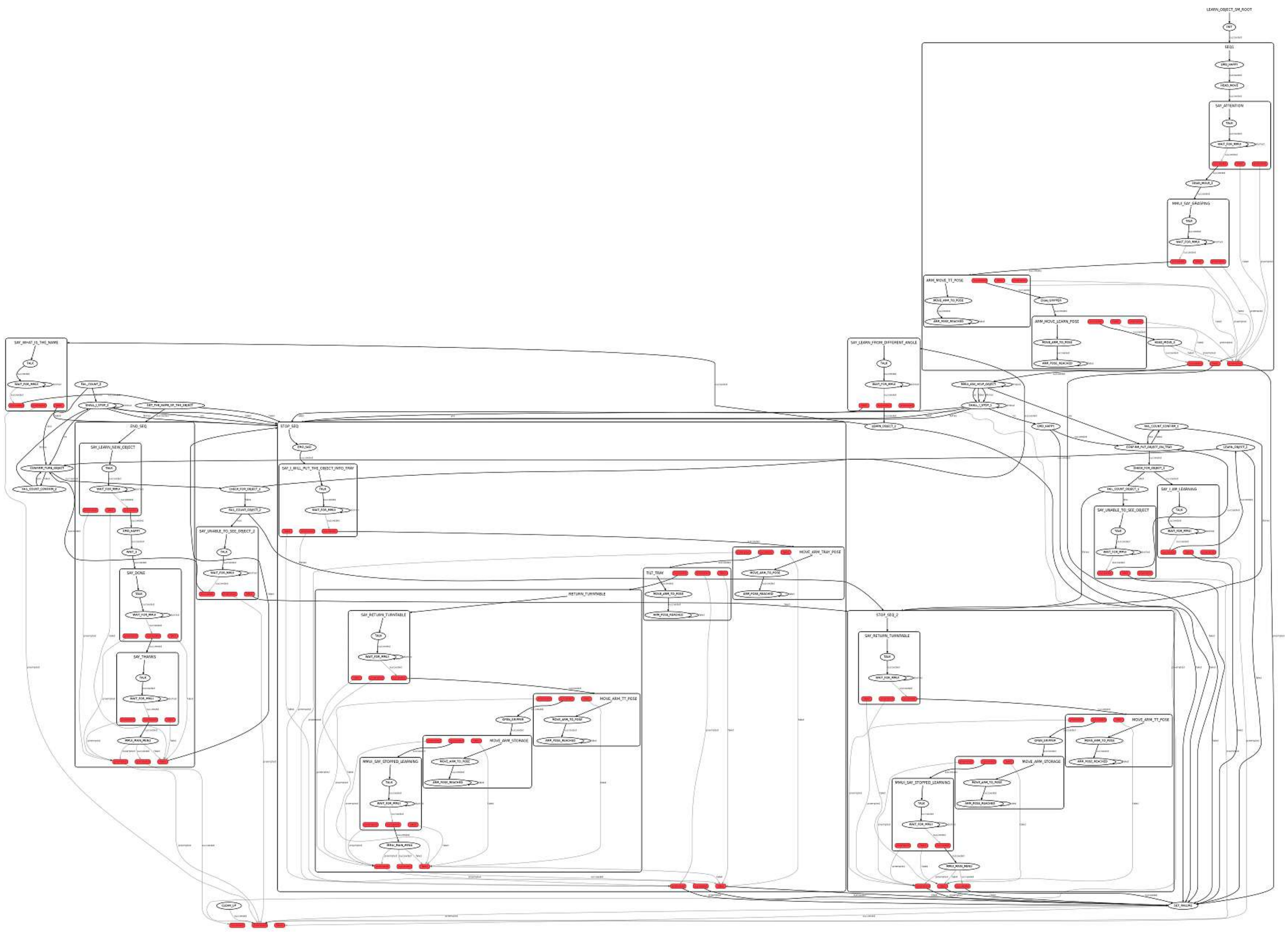


Figure 3.11: The diagram of the state machine of the Learn object task examplifies the increased complexity of Hierarchical Finite State Machines.

Chapter 4

Field Trials of Hobbit's Behavior Coordination System

The main challenge we faced in our research was the scarcity of long-term studies involving robots operating in private households. As the aging population increases in many countries, the importance of using technology, such as service robots, to support the elderly becomes more significant.

One of the key uncertainties was whether a service robot could autonomously function in private homes for extended periods without constant supervision from support personnel. Additionally, we were unsure about the specific functions and features that such a service robot should offer and which functionalities people would actually use, considering the potential decline in novelty over time.

To address these challenges, we took several important steps. First, we developed a cost-effective mobile robot platform. Then, we designed and implemented a behavior coordination system to facilitate smooth interactions between the robot and users. We also created multiple software modules to provide specific functionalities, catering to user needs and enhancing the robot's autonomy. Finally, we integrated all these modules into five distinct robots.

To evaluate the effectiveness of our approach, we conducted rigorous field trials, lasting three weeks each for 18 participants. These trials allowed us to thoroughly assess and validate our efforts.

According to the data and participant interviews, the users engaged with the robot to perform various functions every day during the trial period. However, some functionalities did not work as expected or exhibited unreliability.

Overall, the field trials demonstrated that our robot could operate autonomously in the private homes of participants for an extended time. Interestingly, the participants did not seem to lose interest in the robot throughout the trials, suggesting that the novelty effect persisted for at least three weeks.

4.1 Introduction of Hobbit

The motivation for Hobbit's development was to create a low-cost, social robot to enable older adults to independently live longer in their own homes. The functionality (described in more detail in Subsection 4.1.2) provided by Hobbit is based on requirements given by elderly care professionals and laboratory studies with the first prototype of Hobbit conducted with 49 users in Austria, Greece, and Sweden.

One of multiple reasons for older adults to move into care facilities is the risk of falling and eventually inflicting injuries. To reduce this risk, the "must-haves" for Hobbit were *emergency detection* (detecting a previously fallen person or recognizing the fall of a user if it happens in the field-of-view of the robot's sensors), *emergency handling* (automatic phone calls to relatives or an ambulance service after an emergency was detected), as well as *fall prevention* (searching and bringing known objects to the user and picking up items from the floor pointed to by the user). Hobbit also needed to provide a *safety check* feature that guides the user through possible risks in specific rooms (e.g. wet floor in the bathroom, slippery carpets on wooden floors) and explain how to reduce these risks. Further, potential end-users and care personnel emphasized the importance to provide a *basic fitness program*, designed to enhance the user's overall fitness.

To increase the acceptance of Hobbit [Lammer et al. [67]] introduced the *Mutual Care (MuC)* concept to enrich the expressed behaviors of the machine. MuC's goal is to create an emotional bond between the person and the robot in such a manner, that it doesn't only provide useful assistance to but acts as a companion as well.

To gain insight into the needs of elderly living alone we invited Primary Users (PU), 75 years plus and living alone, and Secondary Users (SU), who are in regular contact with the PU, to workshops in Austria (eight PU and 10 SU) and Sweden (25 PU).

We explored needs, desires, and expectations from end-users by collecting their input at an early stage of the Hobbit project [Körtner et al. [63]]. A questionnaire survey with 113 PU in Austria, Greece and Sweden, as well as qualitative interviews with 38 PU and 18 SU were conducted. This user-centered design approach allowed us to collect a large amount of possible features, of which a subset was realized in the final implementation. Further, this process gained us important knowledge for the aesthetics in terms of the design, material, and motion behaviors for the different design stages of the robot Figure 4.1. To make the best-informed decision on which features to choose further laboratory trials with the first generation of Hobbit (PT1) was performed. The results of these trials [Fischinger et al. [40]] laid the foundation for the final feature list for Hobbit.

1. **Call Hobbit:** Summon the robot to a position linked to one of multiple wire- and battery-less AAL buttons.

Figure 4.1: From left to right: First mock-ups designed by secondary users, the first generation Hobbit prototype (PT1) and Hobbit as used during the field trials.

2. **Emergency:** Automated phone call to relatives or an ambulance service to inform them of an *emergency* situation. Can be triggered by the user from emergency buttons, commands (speech, gesture, or touch) or by the robot (fall recognition or fallen person detection).
3. **Safety check:** Guide the user through a list of common risk sources and provide information on how to reduce those risks.
4. **Pick up objects:** Objects lying on the floor should be picked up by the robot with no distinction between known or unknown objects.
5. **Learn and bring objects:** Visual learning of user's objects to enable the robot to search and find them within the apartment.
6. **Reminders:** Deliver reminders for drinking water and appointments directly to the user.
7. **Transport objects:** Reduce physical stress on the user by placing objects on to the robot and let the robot transport it to a given target location.
8. **Go recharging:** Autonomously or by the user's command the robot moves to its docking station for recharging.
9. **Break:** Put the robot on break when the user leaves the apartment or takes a nap to reduce false alarms when the user is not found during patrol.
10. **Fitness:** Physical exercises that increase the overall fitness of the user.

11. **Entertainment:** Brain training games, e-books, and music.

4.1.1 Robot Platform and Input Capabilities

The Hobbit robot as seen in Figure 4.2, is a custom-built platform with a rectangular footprint (width 0.48 m, length 0.55 m, and height 1.25 m). On its right side, the robot is equipped with a 6-DoF arm and a two-finger gripper, mounted in such a way that objects lying on the floor can be picked up and placed on a tray on top of the robot's body. Furthermore, the arm can grasp a small turntable stored on the right side of the body, which is used by the robot to learn unknown objects.

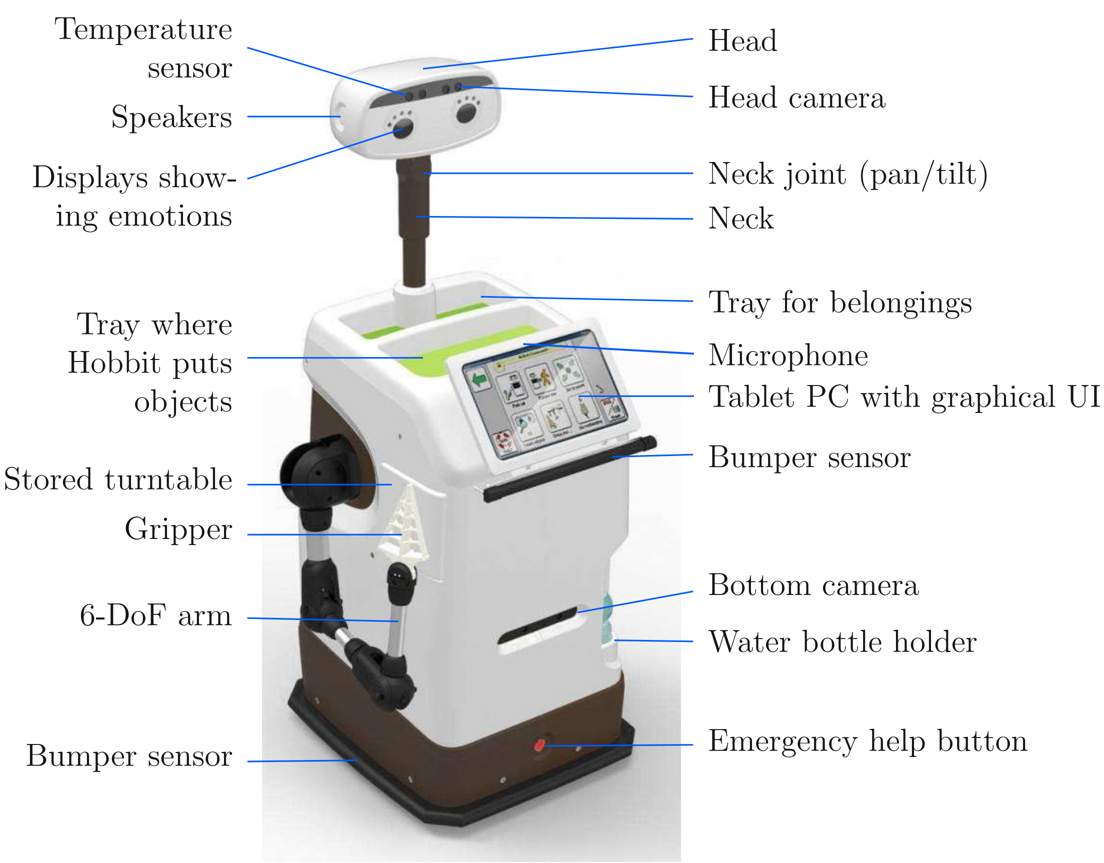


Figure 4.2: Rendering and hardware description of the second prototype of Hobbit (PT2).

On the front of the robot is a tablet computer that provides the Multi-Modal User Interface (MMUI) as shown in Figure 4.3. Generally speaking, the MMUI is a framework containing the following main building blocks: Graphical User Interface (GUI) with touch, Automatic Speech Recognition (ASR), Text To Speech (TTS), and Gesture Recognition Interface (GRI). The MMUI provides the emergency call feature, web services (e.g. weather, news, shared calendar, and social media), control of robotic functions, and entertainment features. Figure 4.4 shows two use-cases of the MMUI. More details about the technical implementation of Hobbit are presented in [Bajones et al. [10]], [Vincze et al. [113]].

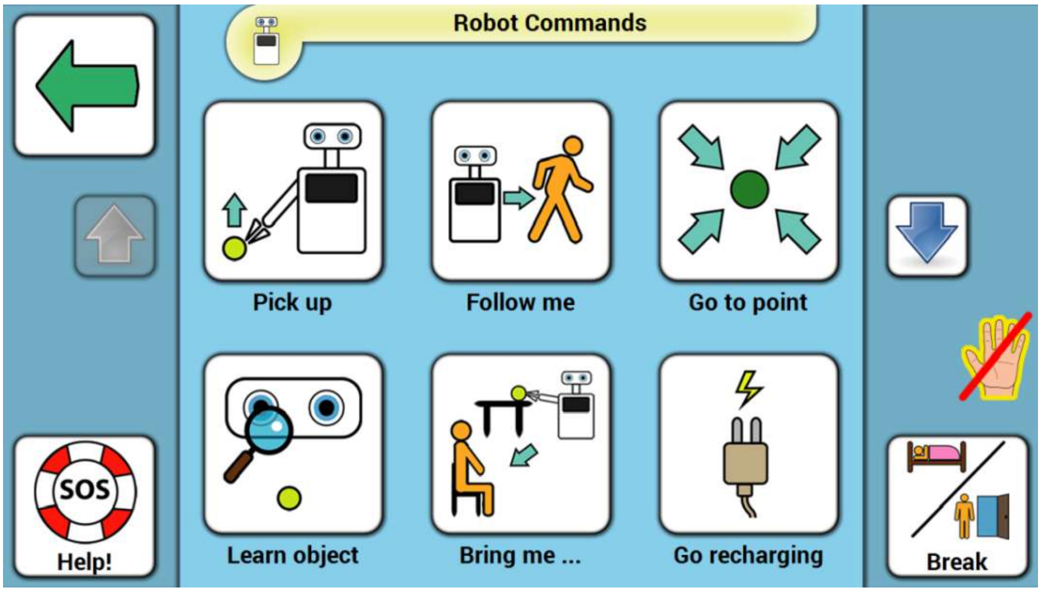


Figure 4.3: Robot command *interface of the Multi-Modal User Interface of Hobbit PT2. The six most often wished for functions are shown in the center of the screen. On the right side an indicator shows that gesture recognition is currently unavailable. On the bottom left the software version of the emergency button (available in every menu of the user interface) is shown; the break button (used to inform the robot that the user is leaving the apartment or going to bed) is shown on the bottom right.* [1]

4.1.2 Autonomous provided behaviors

The implementation of all of Hobbit's behaviors was done in the process of translating the user requirements into scenarios in the form of stories. Each scenario was designed as a flow-chart with well-defined states in which the robot expects input from the user. This does not mean however that input outside these states is not handled, the behavior coordination decided if a running task should continue its execution or if it needs to be prematurely stopped before a new task can be started. This decision-making was based on the priorities of the active and to be started scenario (where *Emergency* has the highest priority). The trigger to start a scenario could come from different sources such as speech, touchscreen, gesture recognition[2],

[1]The symbols for *gesture* or *speech input not available* as well as all other parts of the user interface were designed in corporation with our end user partner, health and elderly care provider *Academy for Aging Research* to ensure that they are easily understandable by our target group.

[2]The trigger from the gesture recognition was ignored under certain circumstances. For example when the robot was looking at the rotating turntable during the *Users teaches Hobbit a new object* scenario.

scheduled events or from a physical button (e.g. emergency help button on the robot or a wireless button in the bathroom). All of the following behaviors were available within the field trials.

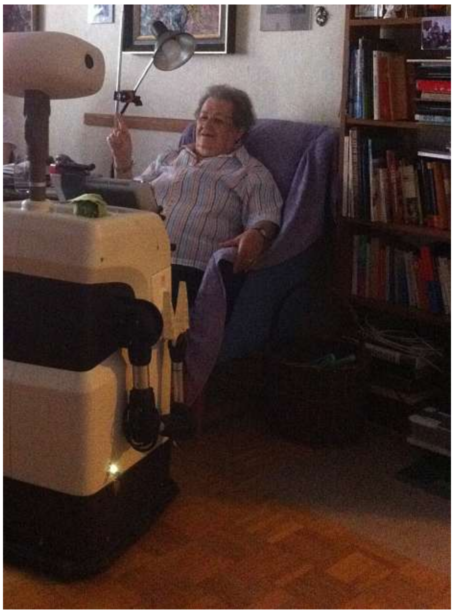
Emergency handling was declared the main and most important functionality of the Hobbit robot. An *emergency* situation can either be declared by the user by

- pressing on the wireless emergency button in the bathroom.
- pressing the physical emergency button at the lower front of the robot (in case of the person lying next to the robot)
- using the touch screen interface on the tablet on Hobbit
- giving a speech or gesture command

It can further be done autonomously by the robot when

- the person detection module detects a fallen person or a person falling in front of the robot
- it is unable to locate the person during the patrol

When the *emergency* situation has been started the robot moves close to the user (only if the user has been detected) to make it easier for the user to reach



(a)

(b)

Figure 4.4: Hobbit provides multi-modal input methods such as simple gestures, touch, and speech input. (a) shows the No *gesture which can be used to answer Yes-or-No questions. In (b) a similar question is answered with* No *on the touch interface.*

for the water bottle. Afterwards the robot calls a relative or emergency contact through the integrated VoIP system. At the start of the call a prerecorded message explaining the situation is played, before both people can talk directly to each other over the microphone and speaker system of Hobbit.

Call Hobbit was the method to command the robot to move to the user's current location. The person would press one of multiple wireless buttons to initiate this task. As each button was associated to a certain location within the environment, the robot would move to the location linked to the specific button that was pressed. Upon arriving at the target location we used the person detection and tracking feature, and gesture recognition to adapt the final location of the robot. This was necessary to follow the fact that our users had different preferences on how close Hobbit should come to them.

Patrol scenario was supposed to make sure that Hobbit was in regular contact with the user. Three hours was decided as the maximum duration to pass since any interaction between the person and Hobbit. This time limit was designed to make sure that a fallen person would be detected by the robot before dehydration of the body shows negative effect. When three hours had passed since the last interaction, Hobbit moved from the current location (i.e. charging station) to multiple way-points until it visited all rooms of the apartment. [3] After Hobbit arrived at a way-point, it would rotate by 360° while performing a person detection. The head-mounted RGB-D sensor was directed slightly downwards to detect an either lying or standing person. When the person was detected Hobbit initiated a conversation that would only succeed when the person gave a reply, so to make sure they are fine. If this interaction did not succeed, or the person was not detected Hobbit started calling out (at full volume) for the user, in a five-minute interval for a total of 15 minutes. The idea behind this was to make sure that it did not miss the person while moving to another room or to capture the attention of a person in a room the robot could not enter. Upon a failure to get a reply in any of these situations Hobbit would initiate the *emergency* handling behavior.

Safety check worked as a way to educate the user about typical, potential risks within their homes. A list of such risks and how to avoid them was collected and agreed on. Such risks typically include a wet floor in the bathroom, a rug that can slide, and an unintentionally turned on stove or oven. On the first day of use, Hobbit encouraged the user to go through this list and confirm that they are aware of them.

User goes away from home or to sleep handled the method of letting Hobbit know that the user is not to be disturbed for a given amount of time (e.g. until next morning or one hour). When selected from the MMUI a checklist *You*

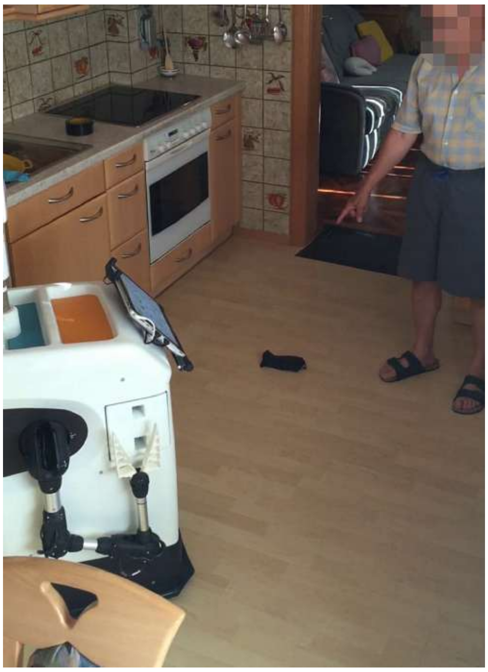
[3] Depending on the environment some rooms were not visited due to privacy or space restrictions.

should check if you turned off the stove. or *Do you want to be reminded of today's appointments?* was presented. Afterwards the robot asked for how long the user wishes the robot not to patrol through the apartment. To use the time when the user was not around Hobbit would then drive into the charging station to make sure the batteries were charger for later use.

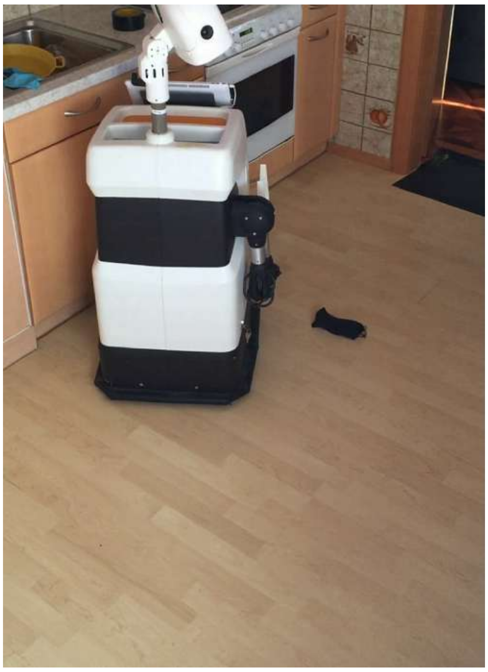
Pick up an object from the floor was the semi-automatic option to remove a desired object from the floor and bring it to the user. By clearing the floor of unwanted items Hobbit was supposed to minimize the risk of a person stepping and slipping on them. Started on the MMUI and pointing with a fully stretched arm towards an object on the floor a person could command Hobbit to pick up a specific item. It would then move to a position from where the item is observable and reachable by the robot and calculates a grasp point and approach trajectory with the HAF algorithm [Fischinger et al. [39]]. After a possible fine-adjustment of Hobbit's position the arm was extended and reached for the object before placing the object on top of the robot on the built-in tray. As the calculations are only based on depth data no prior model or other information of the object is necessary. The full sequence is depicted in Figure 4.5.

User teaches Hobbit a new object for later object search and fetch-and-carry tasks. We aimed at reducing the time a person would need to walk around the apartment while searching for a specific object. As this task requires the robot to have prior information of the desired item it needs to acquire them with the help of the user beforehand. Hobbit would guide the user through the process in a step-by-step fashion and started by retrieving the turntable from its location on the right side of the robot's body. The object was then placed on top of the turntable that rotated while the RGB-D sensor in the head was pointed towards the turntable. A continuous stream of RGB and depth information was stored for later processing in which a full 3D model was assembled [Prankl et al. [84]]. This was done during an idle time in the charging station as to not limit the robot during any other task. The label (necessary for the *Bring object* task) of the item was entered with an on-screen keyboard at the MMUI.

Bring object To help the user find objects in their apartment we equipped Hobbit with this feature. After the data of an item has been recorded and the model was successfully built, this model could be used by the robot to search and find the item. For that Hobbit moved through the apartment to specific way-points, that had been selected to optimize the Field of View (FoV) towards tables, shelves and cupboards. At these locations we used an RGB-D based object recognition and pose estimation [Aldoma et al. [2]] to locate the desired item. When possible the robot grasped it and placed it on top of its tray, if not it would search the user to inform them about the location of the object.



(a) In the first step the user points at the object on the floor while standing in front of Hobbit



(b) The robot moves to a location from which it can better inspect the area on the floor

(c) After the object was detected, and grasping points were calculated, Hobbit grasps the object.

(d) In the final step Hobbit drops the object on to its tray.

Figure 4.5: Hobbit performs the task of picking up a soft spectacle case during the trials for a user in Vienna

Reminders This behavior was a simple way to assist the user in keeping track of medication and appointments, and made sure that they did not forget to drink enough water during the day. The system consisted of two parts. First, the web-interface to add, modify or delete calendar entries for the SU. Second, the scenario in which Hobbit would be triggered before the calendar entry to find the user in the apartment and deliver the reminder. The actual implementation re-used parts of the *Patrol* functionality. Reminders were available in different categories, such as *medication*, *appointments*, and *drinking reminders*.

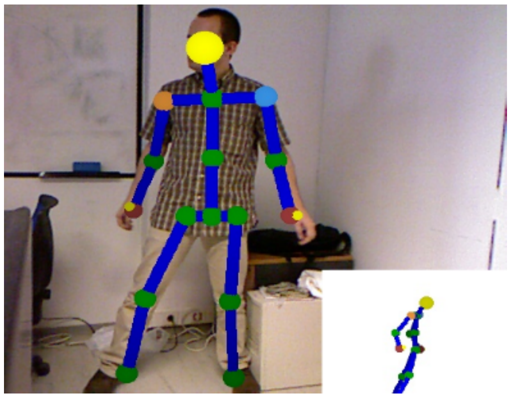
Transport object Carrying heavy items poses an increased risk of falling for elder people. To reduce the risk Hobbit, as a move-able robotic platform, is able to transport these items from one location to another. Two options for this task were developed, both of which started after the user placed the objects into the tray on top of Hobbit. First, a simple *Go to* command, that let the user decide beforehand where the robot should bring the object. Second, used the person tracker to locate the user and plan a continuously updated path behind the person.

Fitness functionality as a means to increase the physical fitness of the PU was a main task considered in Hobbit's development. As a precondition, the user sits on a simple chair with sufficient space surrounding her/him to be able to move freely with the arms. Hobbit moves into a position about 1.5 m away, pointing in the direction of the PU. On the integrated tablet instructional videos were shown to the user on the left side of the screen and a skeletal representation of the tracked user's upper body on the right side as shown in Figure 4.6. The user's movements are tracked and analyzed for correctness. If the exercise was performed correctly, positive feedback was given. If not, instructions on how to correct and improve were presented to the user.

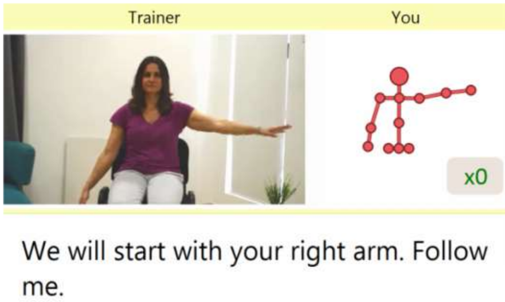
Robot recharging fulfills the technical necessity of the robot to keep a certain battery level to guarantee autonomous operation and can be started by the user or by the robot itself. The charging procedure starts with Hobbit moving to the location of the docking station and detecting its unique shape of it for fine adjustment while driving onto the charging pads. Metal contacts on the bottom of the robot touch these charging pads on the docking station for power transfer. For safety reasons the charging pads are not enabled until the robot identified itself to the station.

Person Detection and Tracking In order to set its behavior appropriately, Hobbit requires to be able to detect people in its vicinity as well as to monitor their body pose. To support this requirement, we built a component that enables Hobbit to detect and track the human body of one or more users based on RGB-D input data from the head-mounted sensors. This component then serves as a building block for developing more competences such as activity

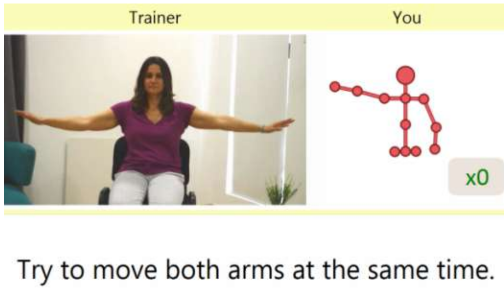
(a) General overview of the fitness setup



(b) Example the full body tracking in form of an simplified skeleton imposed onto the image of a person



(c) Avatar mirroring the trainer's movement proved easier for users to follow.



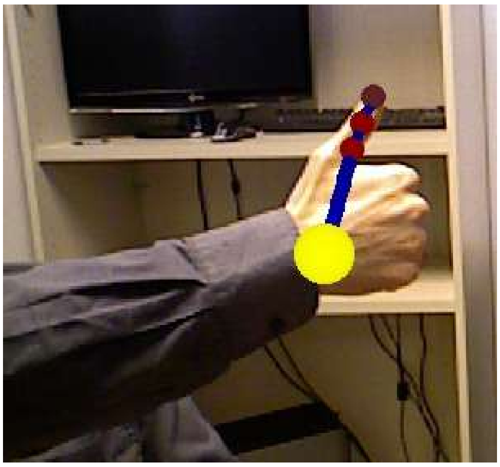
(d) Correction suggested by the system to the user.

Figure 4.6: The top row show how the user is positioned relative to the robot and the fitted and tracked 3D skeleton of the user from the robot's point of view. The bottom row shows examples of the content of the robot's screen that is shown to the user during the fitness exercise. It includes instructions on how to improve based on the detected user's motions.

recognition [Kosmopoulos et al. [65]], natural Human-Robot communication [Michel and Papoutsakis; Michel et al. [76; 77]] as well as specialized functions like the fitness application [Foukarakis et al. [43]]. Due to the fact that elder people spend a large amount of time sitting on a wide ranger of types of chairs, benches, couches or 4-wheeled walker, commonly used skeleton tracking algorithms were not able to reliably detect and track them during early tests. The method we incorporated into Hobbit [Michel and Argyros [75]] provides detection and tracking of a user's body while standing or walking based on frontal, back- or side-views and able to track the 3D position, orientation and full articulation of the human body.

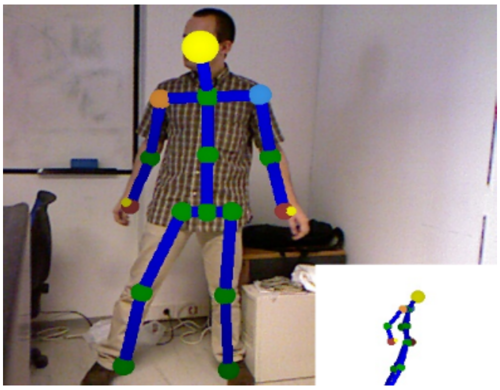
This solution proved to show the following benefits during Hobbit's field trials.

(a) Upper body tracking of a sitting person.



(b) Hand gesture yes *detected.*

(c) Hand and finger detection of a siting person.



(d) Full body tracking of a standing person.

Figure 4.7: Qualitative results of the 3D skeletal model-based person detection and tracking method.

- Non-intrusive, marker-less tracking of the person's body in 3-dimensional space.
- No training data of the user has to be collected.
- Works on a single, inexpensive RGB-D sensor.
- Shows high robustness under challenging and often changing conditions (e.g. camera motion, illumination changes, partial occlusions).
- Automatic recover after loss of tracked person
- Provides a 2D fallback solution based on the well-established Viola-Jones face detection algorithm [Viola and Jones [115]].
- Performs close to realtime on a standard computer in the year 2014.

Qualitative results of the method are illustrated in Figure 4.7.

Fall Detection End-user's need assessment and the results of PT1 laboratory studies showed that a method to find and assist a fallen person was a must-have feature of Hobbit. As the person should receive help, either to get back on their own feet or medical attention, Hobbit needs to reliably detect when she or he fell down. Three possible situations needed to be handled by the robot. First, when the person is falling while in the FoV of Hobbit's top RGB-D sensor. The method to handle this is a rather simple classification algorithm, trained on positions and velocities of tracked bodies joints of a person, given by the former discussed person detection and tracking techniques. When the tracked body falls down to the ground the classifier triggers and the *emergency* procedure starts. Second is the more likely situation, the person falls down while the robot's sensor is not able to observe the fall. In this case the robot must first find the person before it can call for help. The fact that this fall might happen in a room Hobbit was not allowed or able to move into is the third case and demands special care.

The simple application of the person detection and tracking module for the second case poses the problem that typical assumptions of body detection, such as that the head is above the shoulders, no longer necessarily hold true. To work around this issue we used the combined data of both RGB-D and the temperature sensors in the head of Hobbit.

Given the known pose and height of each camera with respect to the platform base, we apply 3D dominant plane estimation in order to detect the floor plane (Figure 4.8). This involves a custom implementation of a RANSAC based plane segmentation algorithm [Strutz [106]]. We compare the orientation of each candidate floor plane to an expected result. Having extracted the floor plane, we segment any outlying objects using depth-based connected components, and we focus on the dominant object in the center of the field of view.

The dominant object, if any, is analyzed and classified based on its volumetric characteristics as a lying human or an obstacle. We also take advantage of temperature measurements acquired by the thermal sensor mounted on the robot's head and calibrated with the head depth sensor. A single temperature measurement is provided for the central pixel of each acquired depth frame. Multiple heat readings are acquired on the surface of the detected blob and in its surrounding area by moving the robotic head, thus the optical axis of the thermal sensor, appropriately.

Temperature values are collected and evaluated (see Figure 4.9). A temperature measurement of the central pixels exceeding 29-35°C, indicates a possible human body in the FoV of the sensor. Finally, given the extracted volumetric characteristics of the detected object from both the cameras' viewpoints and the temperature measurement, a classifier is used to compute the final result among the categories of a lying human, an object or empty space. The

Figure 4.8: Lying user detection during lab tests. From left to right the acquired pair of RGB-D frames that forms the input to the plane estimation method. Two rightmost frames: plane detection and removal, and dominant outlying object detection.

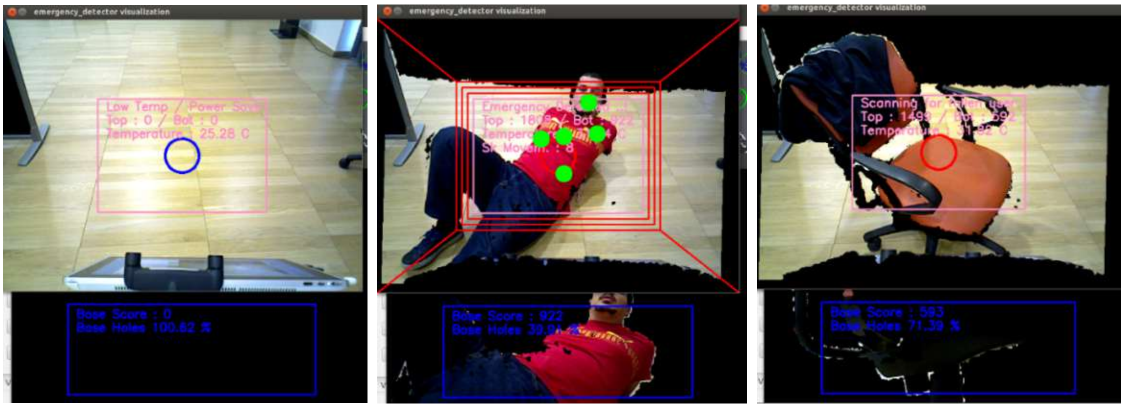


Figure 4.9: Vision-based emergency detection of a fallen user lying on the floor. The upper and lower middle images show the captured frame from the head and bottom cameras, respectively. The green dots mark a found skeleton within the search area (green and blue rectangles). From left to right: No human, no detection; person lying on the floor, correct detection; volumetric data from the head's depth and temperature sensor are in conflict with the volumetric date provided by the bottom depth sensor.

detected object is also checked against the Hobbit global map to exclude mapped obstacles that may have a human-like temperature and comparable volume to a human body (i.e. radiator, electronic devices etc.). The fall detection component was always actively running in a separate thread to guarantee the detection of a fallen or falling person under any circumstances.

Gesture Recognition Besides the main input modality of touch input on Hobbit's tablet speech and gesture recognition was incorporated. While for speech recognition off-the-shelf ASR technology was purchased, the gesture input module was developed within the Hobbit project. The intent behind this was to give the user a second method of interaction when the robot is not within reach or when ASR is not working well because of the ambient noise (e.g. radio or TV). As Hobbit already included a person detection and tracking module the extension to recognize certain gestures was a logical improvement to the robot. However, we did limit the amount and complexity of the gestures so that the duration of necessary training of our users could be reduced to only a few minutes. A further difficulty lied within the cultural differences of our user base as gestures are not as universal as generally believed. The possible gestures we agreed on and implemented are shown in Figure 4.10 and explained in detail in Table 4.1.

Our framework for gesture recognition [Michel et al. [77]], [Michel and Argyros [75]] consists of a complete system that detects and tracks arms, hands and fingers and performs spatiotemporal segmentation and recognition of the set of predefined gestures, based on RGB-D data from the head sensor of Hobbit. Once the palm and fingers have been detected, their pose is estimated and their trajectories over time is analyzed to classify them into one of the possible gestures or into an unknown movement. A thorough evaluation of the gesture recognition with elder users has been previously reported [Michel et al. [77]].

Figure 4.10: Integrated human body detection and tracking and gesture recognition. The output of the former module feeds the gesture recognition method based on 3D position and angles of the tracked body joints. Left: "Come closer" gesture of a standing person. Right image: "Yes" gesture of a sitting user. Lower image was previously reported in [Michel et al. [77]]

Mutual Care as Underlying Interaction Paradigm

To investigate the impact of the MuC concept we implemented two different social roles that Hobbit can fall into. These social roles are the *device* mode, in which MuC

Table 4.1: Set of hand/arm gestures recognized by Hobbit

User Command	Upper body gesture	Robot command	Related Scenarios/ Tasks
Yes	Thumb up-palm closed	Positive response to confirmation dialogues. YES gesture.	All (1 m to 2 m distance to robot)
No	Close palm, waving with index finger up	Negative response to confirmation dialogues	All (1 m to 2 m distance to robot)
Come closer	Bend the elbow of one arm repeatedly towards the platform and the body	Reposition the platform closer to the sitting user	All (1 m to 2 m distance to robot)
Cancel task	Both open palms towards the robot	Terminate an ongoing robot task	All
Pointing	Extend one arm and point in 3D space towards an object (lying on the floor)	Detect & grasp the object of interest towards the pointing 3D direction	Pick up an (unknown) object from the floor
Reward	Open palm facing towards the robot and circular movement (at least one complete circle is needed)	Rewards the robot for an accomplished task	Approach the user
Emergency	Cross hands pose (normal- range interaction)	Emergency detection, initiated by the user	Emergency detection

was completely disabled, and *companion* mode. In the latter, Hobbit is proactively engaging the user, adapting the distance between itself and the user, stay longer in the same room as the user after a task was completed and used a more intimate style during the dialogues. Additionally, the feature *return of favour* was enabled in the *companion* mode. These changes manifested themselves as

Return of favour Hobbit asked if it could return the favour when the user had assisted Hobbit during a task.

Communication style Hobbit used the user's name in the dialogue and was more human-like such as responding to a reward from the user by saying *You are welcome* instead of *Reward has been received.*

Pro-activity Hobbit initiated interactions instead of waiting for a command from the user or from a triggered event such as a calendar entry.

Presence The robot stayed with the user even when the current task has been finished. The user could either send Hobbit away explicitly or wait for up to 30 minutes until it would move back to the charging station.

The switch from *device* to *companion* mode happened at the half of the field trial of the current user (day 11) without any explanation by the robot or anybody else to reduce the risk of anticipating a certain behavior.

4.2 Field Trials

We conducted field trials in the households of 18 PU with 5 Hobbit robots in Austria, Greece, and Sweden. The trials lasted ~21 days for each household, resulting in 371 days overall. During this time the robots were placed in the homes of 18 older adults living on their own, where users could use and explore the robot on a 24/7 basis.

The trial sample consisted of 16 female and 2 male PU, their age ranged from 75 to 90 years (M=79.67). All PU were living alone, either in flats (13 participants) or in houses. In adherence with inclusion criteria set by the research consortium, all participants had fallen in the last two years or were worried about falling and had moderate impairments in at least one of the areas of mobility, vision and hearing. 15 PU had some form of multiple impairments. Furthermore, all participants had sufficient mental capacity to understand the project and give consent. In terms of technology experience, 50.0% of the PU stated to be using a computer every day, 44.45% stated to be using a computer never or less than once a week and only one participant used a computer two to three times a week.

Exploring HRI with Older Adults It was shown in Human-Computer Interaction research that older people tend to praise the developers rather than giving an objective view, thereby being very positive about prototypes they are presented and tending to blame themselves rather than the interaction modalities if not being able to cope with the system [Eisma et al. [34]]. Age-related factors can also make self-reporting inaccurate (for example, in questionnaires), with recent research showing that there are differences due to age in the way in which people respond in self-reports, which we intended as an indication for a multifactor method mix for the long-term household evaluation [Marquié et al. [73]]. A good example of these effects was observed in our first laboratory user study. In the final interviews of the study, some users considered themselves as too healthy and active to need a Hobbit at home. We consider this partly as an answer effect, as it would be stigmatizing for an older adult to admit that they need a Hobbit to independently live at home. Moreover, clearly the prototype level will have impacted this reply as it might be difficult for an older adult to imagine the full capacity of Hobbit at the next prototype stage. Subsequently, we decided to use a multifactor method mix and a multi-informant approach for the field trials combining self-reporting data with logging data to gather more objective data about the usage of Hobbit and to involve also secondary users (i.e. friends and relatives).

Field studies So-called evaluations "in the wild" are studies [Rogers [88]] where researchers are decamping from their labs and moving into actual users' environments (e.g. homes), which is often claimed to be of utmost importance for HRI research [Sabanovic et al. [92]]. A central part of evaluating outside the lab involves observing and recording what people do and how this changes over suitable periods of time. Whereas the burning question in HRI studies used to be "How many participants do I need?" the hotly debated question now is "How long should my study run for?"[Rosenthal-von der Pütten et al. [91]]. Some researchers argue for a few weeks, others suggest months or even years are needed to show sustainable and long-term effects. However, the most crucial question is: "How can we measure what we really need to know without having a researcher present?" and "How can we conclude causality from statistical correlations without being present?".

Exploring Various Factors Over Time It is well-known that the first impressions of robots are most often positive (so-called encounters at the zero acquaintance level) and to study these impressions is valuable for HRI scenarios [Dautenhahn [28]] where the application context requires brief non-repeated interaction (e.g. a museum guide robot). However, this is definitely not the case for an elderly care robot at home. Preferences and attitudes of users are likely to change over time and novelty effects will wear off. Carrying out studies over several weeks in the private homes, which take into account several quality factors are labor-, time- and equipment-intensive, but require the understanding of how a robotic product may become a social product.

Several long-term studies in HRI choose the approach to study only specific functionalities at specific phases of the study, e.g. first week one function, second week another function etc. (see e.g. [Leite et al. [70]]). This approach should guarantee that the participants do not experience a cognitive overload in the beginning of a long-term study. However, we intentionally decided against such an approach, as it is our overall goal to assess the holistic experience of Hobbit as a product. As soon as Hobbit becomes a real product it will also offer all its functionalities at once and there will only be an initialization phase in which the user becomes acquainted with the robot. Therefore, we need the field trials as an ecological valid reference to identify: Which functionalities are actually used by the participants over a longer period of time, and how does the interaction change after the novelty effect is over? A reasonable benchmark can, to our conviction, only be achieved if the robot offers the same interaction spectrum over three weeks.

The fact that Hobbit should autonomously interact with the older adults and also manipulate objects in their homes brought along challenges in terms of a controlled methodological approach. Primarily, we had to ensure that users were able to interact with the robot on their own, without an influencing observer constantly present. To moderate these difficulties, multiple methods were applied to gather insightful data on usability and acceptance. We chose a combination of qualitative measures and quantitative measures (attitudinal as well as behavioral). Data was collected from multiple sources ("multi-informant approach"): data of the PU, interview data from the SU, and data logged by the robot itself.

User behavior was assessed several times during the user trial and continuously tracked through the robot's logging data, respectively. All participants experienced both roles of the robot: "device-like" and "companion" (Mutual Care). For the first 11 days of each trial, the robot was set to "device-mode", afterwards it was changed to "companion-mode". In order to avoid a bias brought about by expectations or information that could prime the users, participants were not told about this change and its effects beforehand. Due to the small sample size, we decided for this within-design in which all participants experienced the robot in the same fashion. Quantitative and qualitative data collection was scheduled in accordance with the change of the MuC mode, i.e. the *Midtrial* assessment allowed for collecting baseline data of the period in which the PU experienced Hobbit for 11 days with all features except MuC. After that, MuC "companion-mode" was switched on and users were surveyed at the *End-of-trial* assessment again. Changes in user acceptance between *Midtrial* and *End-of-trial* are therefore mainly due to the MuC feature. Additional *Pre-phase* and *Post-phase* assessments complemented the data source and allowed for changes to be clarified over time.

4.2.1 Questionnaires/Scales

For the quantitative assessment of the interaction with Hobbit, we aimed at using existing, validated scales as much as possible, as the sample size of the trial is rather small. However, for the acceptance indicators emotional attachment and perceived reciprocity, we needed to generate our own items. Moreover, feedback from PU and SU showed that even already existing scales had to be adjusted for the users in terms of wording or answering format.

Questionnaires used in the trials were: (1) Falls efficacy scale, (2) Self-efficacy scale, (3) Negative Attitudes Towards Robots-scale (NARS), and (4) Self developed items on emotional attachment and perceived reciprocity.

The Falls-efficacy Scale [Yardley et al. [125]] measures fear of falling by looking at self-reported difficulties with physical and social activities. For the purposes of our trials, a shortened version was used, which consisted only of the items that deal with fear of falling in the home.

The Self-efficacy scale [Schwarzer and Jerusalem [96]] is an established questionnaire to measure to what extent a person has optimistic self-beliefs to cope and solve problems (in everyday life) on his or her own. This psychometric instrument is available in validated versions in various languages, including German, Greek, and Swedish. The idea of using this questionnaire was to monitor if using the robot could have a positive influence on the PU's self-efficacy, which then could be interpreted as a motivating effect as well as increase in quality of life and increase of feeling independent at home.

The NARS [Nomura et al. [81]] serves the purpose of measuring negative attitudes towards robots. We expected that the attitude of users will become more positive (or at least not become more negative) in the course of the trials. NARS exists in different non-standardized translations. In order to render the single items easier to understand for the Hobbit trials, wording was changed where necessary and formulated in a more neutral way.

In order to study the acceptance indicators emotional attachment and perceived reciprocity between PU and robot, we generated some specific items, asking users whether they felt the robot was like a companion or a pet, whether they had the feeling that the robot had its own will etc.

Please refer to the Appendix for precise details regarding the interview guidelines A, timeline B and questionnaires C, D, E, F, G, and H.

4.2.2 Interview Guidelines

The in-depth interviews should serve the purpose to learn about individual experiences and opinions, as well as an opportunity for the participants to reflect on their interaction with Hobbit over time. Topics covered in the first interview with the PU were the perception of the robot's behavior and how the user would describe the robot. Topics covered in the second interview were usability aspects (utility, flexibility and ease of learning) and user acceptance aspects (perceived safety, self-efficacy, emotional attachment and perceived reciprocity). Similarly, in the last interview topics covered were utility, perceived safety, emotional attachment and perceived reciprocity. SU were interviewed at the end of week three. Their interviews covered the following aspects: perceived utility of the robot, PU's self-efficacy, and perceived safety.

During the times of the visits, the involved researchers documented how the users interacted with Hobbit. Each interview lasted between 60 and 90 minutes. The interviews were recorded, listened to repeatedly and summaries of each interview were translated into English and sent to a project partner for further analysis. NVivo (a software for qualitative analysis) was used to search for patterns, themes and contradictions within one household and between all the households. The interpretations were cross-examined by all the involved researchers and the results presented in this journal paper have been cooperatively generated. The qualitative analysis was based on much more material than the reader will have access to by reading this report. The selected quotes primarily serve as illustrations and have been selected by how they illustrate the most important findings. All questions can be found in the interview guidelines in the Appendix.

4.2.3 Logging Data

To obtain a clearer picture of how often the PU interacted with Hobbit and how these interactions took place we recorded timestamped log data on the internal state of the robot and its behavior coordination system. Further input and its modality from the PU, as well as preprocessed input from sub-systems such as the fallen-user detection, gesture recognition or calendar-based events were saved to the internal logging system. In the trial analysis, the focus was on the following data on usability and user acceptance indicators:

Flexibility Logging the number of times different modalities were successfully used.

Utility Logging the number of robot commands issued by the user.

We assumed that one input modality will be preferred more than others and that specific commands will suit better for one input modality than the other. Data presented in Table 4.2, Table 4.3 and Table 4.4 outline tendencies with respect to preferred interaction modalities. On the first glance it seems that gestures were the most preferred input modality, even more preferred than the touchscreen which we assumed to have the highest usage rate. This counter-intuitive result however, was caused by many false/positive gesture recognitions that happened when the user was moving the arm towards the touchscreen. This was above all true for the over sensitive *come closer*-gesture. This also explains the high numbers for the *come closer*-gesture during most trials. However, content-wise it is convincing that *come closer* and *yes* are the most frequently used gestures as both were used when the robot asked for confirmation to move closer to the user (the most often used action). Similarly, we can argue that *yes* is the most often used speech command, while for touch input the *go to* command and the *recharge* command were most commonly started. Taking into account the many false positives in gesture recognition for the first few participants of the trials, it is rather the case that the touchscreen was used most, followed by speech and gesture.

Table 4.2: Usage Frequencies of Modalities - Austrian users

Modality	AUT1	AUT2	AUT3	AUT4	AUT5	AUT6	AUT7
Gestures	744	293	182	434	31	111	199
Most often	Come	Come	Come	Come	Yes	Yes	Yes
used	434	232	127	323	17	44	88
Speech	476	353	523	78	478	279	344
Most often	Yes	Yes	Yes	Yes	Yes	Yes	Yes
used	307	205	432	34	373	197	166
Touchscreen	473	428	341	216	526	885	443
Most often	Go to	Recharge	Recharge	Go to	Go to/Reward	Go to	Stop
used	104	97	86	51	126	213	119

4.2.4 Pilot Trials

Two pilot trials with older users were conducted with end users in Vienna, Austria. Those pilot trials yielded feedback for the planned evaluation methodology and also showed further challenges and issues to be taken into consideration when bringing a robot into a real home. This included how to transport the robot, what to look for in the set-up phase, what to take care of in the user's home etc. The first took place in July 2014 in Vienna. The individual scenarios were tested with Hobbit in the home environment and technical errors, as well as script failures were logged. Another pilot test took place in September 2014 in Vienna and lasted for two weeks.

Table 4.3: Usage Frequencies of Modalities - Swedish users

Modality	SWE1	SWE2	SWE3	SWE4	SWE5	SWE6	SWE7
Gestures	456	227	401	4	337	23	73
Most often	Come	Come	Come	Fall	Come	Come	Come
used	354	131	336	3	287	17	40
Speech	228	87	167	4	10	40	56
Most often	Yes	Yes	Stop	Yes	Yes/Stop	Stop	Stop
used	119	49	53	3	3	19	20
Touchscreen	689	286	487	7	122	91	492
Most often	Go to	Go to	Recharge	Social Role[4]	Go to	Recharge	Go to
used	194	68	114	4	44	24	233

Table 4.4: Usage Frequencies of Modalities - Greek users

Modality	GRC1	GRC2	GRC3	GRC4
Gestures	247	237	20	12
Most often	Come	Come	Cancel	No
used	229	191	10	5
Speech	34	47	19	6
Most often	Yes	Yes	Stop	Yes
used	13	13	13	5
Touchscreen	162	155	58	10
Most often	Go to	Recharge	Recharge	Recharge
used	27	47	36	8

Table 4.5: Usage Frequencies of Robot actions - Austrian users started/canceled

Action	AUT1	AUT2	AUT3	AUT4	AUT5	AUT6	AUT7
Call HOBBIT	340/190	172/81	74/21	97/75	71/42	146/60	349/263
Pick-up	26/24	78/67	18/18	4/2	18/15	272/250	39/35
Teach object	14/1	21/10	8/0	2/1	1/0	0/0	3/2
Bring object	0/0	0/0	0/0	0/0	15/3	36/1	5/4
Follow me	7/4	25/15	3/3	5/5	26/20	18/6	14/7
Go to	89/33	29/8	18/5	46/26	117/24	386/85	41/17
Recharge	49	186	126	16	88	312	168

Functionalities were tested again, as well as autonomous behavior of Hobbit in a real life environment and communication of a real user with Hobbit through the MMUI. All evaluation data were tested and then edited based on the received user

Table 4.6: Usage Frequencies of Robot actions - Swedish users started/canceled

Action	SWE1	SWE2	SWE3	SWE4	SWE5	SWE6	SWE7
Call HOBBIT	242/123	97/57	340/131	6/4	40/26	39/12	140/61
Pick-up	30/30	19/19	42/40	0/0	9/9	0/0	9/9
Teach object	14/4	4/2	1/0	0/0	1/0	0/0	0/0
Bring object	1/1	2/1	0/0	0/0	0/0	0/0	0/0
Follow me	4/1	7/7	4/2	0/0	0/0	0/0	0/0
Go to	193/109	66/31	84/39	0/0	43/25	18/11	224/70
Recharge	135	46	207	1	0	34	320

Table 4.7: Usage Frequencies of Robot actions - Greek users started/canceled

Action	GRC1	GRC2	GRC3	GRC4
Call HOBBIT	143/49	91/40	50/12	21/14
Pick-up	8/6	10/10	10/9	0/0
Teach object	5/3	1/0	1/1	0/0
Bring object	0/0	0/0	0/0	0/0
Follow me	0/0	4/2	0/0	0/0
Go to	25/13	23/7	9/2	3/0
Recharge	41	78	36	8

feedback. Functional limitations of the prototype identified during the two weeks pilot test were used to trigger further development and system integration. Frequent lab-trials with feedback from researchers were continued until the start of the actual trials to make the system more robust and improve identified faults.

4.2.5 Evaluation strategy

Usability is defined as the overall ease of use; meaning how easily the user can interact with the robot. For the Hobbit field trial, the evaluation of the following usability indicators was considered as relevant: ease of learning, flexibility, and utility. The following research questions were investigated

Ease of learning Is Hobbit intuitive to interact with for the older adults?

Flexibility Does Hobbit provide the relevant input modalities that enable the user to command it effectively?

Utility Does Hobbit offer the right functionalities, therefore older adults feel supported to maintain independent living at home?

User acceptance is in general defined as “the demonstrable willingness within a user group to employ technology for the task it is designed to support” [Dillon [33]].

In order to investigate this multi-layered aspect of Hobbit, a variety of different indicators needed to be assessed, namely: attitude towards robots, perceived safety, self-efficacy, emotional attachment, and perceived reciprocity. The following research questions were investigated:

Attitude towards robots Does the general attitude towards robots change over time due to the interaction with Hobbit?

Perceived safety Does the perceived safety change over time due to having Hobbit in the household?

Self-efficacy Is the self-efficacy maintained on the same level due to having Hobbit in the household?

Emotional attachment Do the users develop an emotional bond with the Hobbit robot over time?

(Perceived) reciprocity Do the users perceive the interaction with Hobbit as a give-and-take relationship?

Representative target users for the trials were older adults who were still relatively independent, but already in need of some support and at high risk of severe outcomes if emergencies are not recognized in due time. Given that the sample for our field study can be considered small from a statistical point of view (18 participants was the ultimate possible number with respect to project resources, i.e. robot platforms built and time), a careful selection of the study sample was crucial. Tightly specifying the participant group provides more reliability given the small group size. Therefore, PU were defined according to the following inclusion and exclusion criteria:

Inclusion criteria:

- Men and women aged 75 or older
- Single-living at home (due to considerations that acceptance of a robot system among senior couples might be lower than for single-living persons).
- At least moderate impairments in one or several of the following areas: mobility or motor skills, sight, and hearing
- Possible multi morbidity (different impairments)
- Possibly also receiving (moderate) home care; help in the household
- Sufficient mental capacity to understand the project and ability to give consent
- Possibility of having an Internet connection in the home (at least mobile Internet)

- Preferably, having a SU such as a relative, friend or neighbor willing to accompany the PU and let Hobbit call them for help if needed

Exclusion criteria:

- Bedridden person/complete immobility
- Any chronic disease that demands regular and longer treatment in hospitals
- Blindness or deafness or unable to interact according to user screening procedure
- Home is unsuitable for Hobbit to be autonomously tested (e.g. stairs, two storied homes, high doorsteps or other barriers, carpeted flooring (carpet not removable), Internet connection not possible to install (e.g. no signal for mobile Internet and no cable-based solution available))
- 24-hour-home-care/regular medical home care that serves to avoid or at least shorten hospitalization of a person
- Cognitive impairments that render the use and understanding of Hobbit impossible (e.g. progressed Alzheimer's disease)
- Pacemaker
- Pets moving freely in the home

Secondary users, such as friends or relatives, were also invited to take part in the trials as support for the PU.

4.2.6 Trial Procedure

All trials took place in private homes of senior adults in Austria, Greece, and Sweden between April and August 2015.

Screening of the users and their homes

Before the actual trials, the homes of potential users were examined to make sure that they matched the inclusion criteria and to discuss possible necessary changes to the home environment for the trials. Most notably, carpets sometimes had to be removed, as well as mirrors. Thresholds between rooms could sometimes be overcome by using ramps. Other obstacles that had to be removed for the trials were usually cables and wires or pieces of furniture in narrow spaces. Internet was made available where necessary. If a user's home fitted the criteria, potential users were screened for the inclusion and exclusion criteria: difficulties with sight, hearing and mobility were assessed in a screening questionnaire. Potential users were also asked if they had fallen in the last two years at home and if they were worried about falling. The higher the participants scored on these scales, the lower their confidence in managing activities of their daily routine. The ability to use the robot independent of sight and hearing difficulties was additionally tested with a more practical approach during the screening-phase: A tablet with the Hobbit MMUI was shown to potential users to test if they were able to read the display from the real distance. Similarly, speech communication was simulated to assess if the potential users were able to communicate via the speech communication tool. Additionally, experience with computers and mobile phones was assessed beforehand. Based on the results and reported difficulties, users were either included or excluded.

Three Weeks Trial

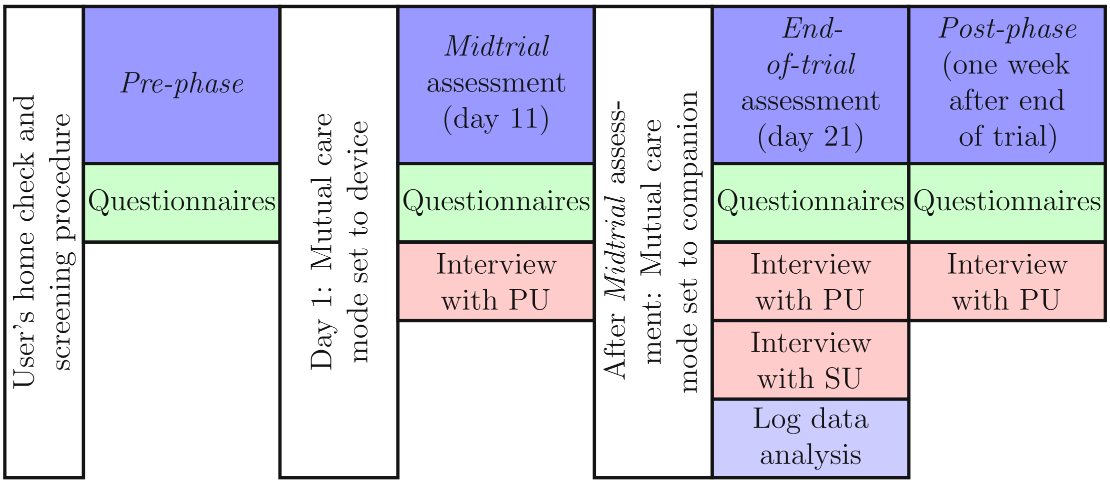


Figure 4.11: Overview of the trial procedure

The actual trial lasted for three weeks for every participating user and consisted of the following phases (for an overview see Figure 4.11). During introduction and *Pre-phase* assessment the robot was brought into the home. Technical set-up took place and users signed an informed consent. The PU completed the *Pre-phase* assessment questionnaires and were instructed in the use of Hobbit. They also received a manual on how to use the robot. Technical set-up and a step-by-step introduction of the functions of the robot took 2-4 days. Afterwards, the user was able to interact with the robot on his/her own. At the *Midtrial* assessment (about day 11 of the trial) users were visited again to complete questionnaires and answer a qualitative interview, then the robot's social role was switched from "device like" to Mutual Care "companion" mode. At the *End-of-trial* assessment (last day of the trial, i.e. about day 21) users were interviewed and completed questionnaires once more. In addition, SU (if available) were also interviewed at the end of the trial. In the *Post-phase* assessment, approximately one week after the trial had ended, PU were visited again for *Post-phase* questionnaires and a short interview. Moreover, log data was automatically recorded by the robot during the whole trial duration. The field trial methodology is comparable to similar studies e.g. [Palumbo et al. [82]].

Trial Support

Each user was provided with a contact telephone number in case of questions and problems they could call any time during the trial. It was consciously taken care to provide PU with one fixed contact for the whole trial instead of several members of the consortium. Technical partners provided the user partners with a list of responsibilities depending on the technical issues in question. This made communication simpler in case of technical issues. Depending on what problem was reported by users, the contact researcher could refer to the list and quickly contact the responsible technical partner for quick troubleshooting. In case of more complex issues, visits by technical partners to the user's home were arranged in order to fix hardware malfunctions.

4.3 Technical reliability and Usage

Technical malfunctions can potentially negatively impact the evolution of a system in a field trial. Therefore, we tried to minimize the risk by informing our users that a prototype of a robot is a very complex technical system which might malfunction. Additional, they were given the phone number of the facilitator who was available for them around the clock, 7 days per week for immediate support. However, malfunctions certainly had an influence on subjects' answers during the assessments and may have attracted attention with the result that subtle behavioral changes introduced by the switch from *device-mode* to *companion-mode* may have been shifted out of the focus. Table 4.8 gives an overview of the functional status across all PU during the field trials. It is based on the combination of i) a check of the robot's features by the facilitator during the pre-, midtrial-, and end-of-trial-assessments, ii) protocols of the calls of the users because they experienced a problem with the robot and iii) analysis of the log data by technical partners. Availability of commands was equally distributed across the two phases of *Mutual Care* (Table 4.8). Please note that an unequal distribution, i.e. that functions unavailable or malfunctioning during *device mode* and then available during *companion mode*, or vice versa, would have introduced additional bias to the evaluation.

The Hobbit field trials marked the first time an autonomous, multi-functional service robot, able to manipulate objects, was put into the domestic environment of older adults for a duration of multiple weeks. Our field trials provided insight on how the elderly used the Hobbit robot and which functionalities they deemed useful for themselves and how the robot influenced their daily life. Furthermore, we could show that it is in principle feasible to support elderly with a low-cost, autonomous service robot controlled by a rather simple behavior coordination system.

Table 4.8: System reliability across 18 PU

MuC mode	Statistics	Call Hobbit	Come closer	Stop Hobbit	Emergency	Pick up object	Teach a new object	Bring object to user	Calendar reminders	Follow me	Move to location
Device	Days total	226	226	226	226	226	226	226	226	226	226
	Days introduction	31	31	31	31	31	31	31	31	31	31
	Days switched off	55	55	55	55	55	55	55	55	55	55
	Days in use	140	140	140	140	140	140	140	140	140	140
	Days when feature was not working	14	13	13	19	84	12	79	49	105	15
	Days when feature only partially working	23	20	22	44	47	116	62	83	13	32
Companion	Days total	148	148	148	148	148	148	148	148	148	148
	Days switched off	20	20	20	20	20	20	20	20	20	20
	Days in use	128	128	128	128	128	128	128	128	128	128
	Days when feature was not working	14	10	8	12	83	17	85	54	92	9
	Days when feature only partially working	25	16	17	38	40	95	43	64	18	29
Device	feature fully working over days in use	81.79%	83.57%	82.86%	70.71%	23.21%	50.00%	21.43%	35.36%	20.36%	77.86%
Companion	feature fully working over days in use	79.30%	85.94%	87.11%	75.78%	19.53%	49.61%	16.80%	32.81%	21.09%	81.64%

MuC mode	Statistics	Go recharge	Take a break	Telephone	Information	Surprise me	Entertainment Audio	Entertainment Games	Entertainment Fitness	Reward
Device	Days in total	226	226	226	226	226	226	226	226	226
	Days of introduction	31	31	31	31	31	31	31	31	31
	Days feature was disabled	55	55	55	55	55	55	55	55	55
	Days of feature in use	140	140	140	140	140	140	140	140	140
	Days when feature was not working	19	16	19	11	11	11	23	20	11
	Days when feature only partially working	20	6	22	9	23	7	20	27	6
Companion	Days in total	148	148	148	148	148	148	148	148	148
	Days feature was disabled	20	20	20	20	20	20	20	20	20
	Days of feature in use	128	128	128	128	128	128	128	128	128
	Days when feature was not working	10	8	22	8	8	7	26	19	14
	Days when feature only partially working	33	9	16	7	23	8	11	23	7
Device	feature fully working over days in use	79.29%	86.43%	78.57%	88.93%	83.93%	89.64%	76.43%	76.07%	90.00%
Companion	feature fully working over days in use	79.30%	90.23%	76.56%	91.02%	84.77%	91.41%	75.39%	76.17%	86.33%

4.4 Trial results

We report the data from 16 participating primary users. Broken down by site, the trials were distributed as follows: Vienna (Austria): seven PU and six SU, Heraklion (Greece): four PU and four SU, and Lund (Sweden): seven PU and six SU. The age of PU ranged from 75 to 89 years with the average age of the sample being 79.75 years. The majority (n=14, 87.5%) of participants were female. In regard to the living situation, 68.8% of the sample (11 users) were living in an apartment, and the others in a house. Computer literacy (assessed in the screening questionnaire by the question "How often do you use a computer?") of the sample was clearly divided between either "never / less than once a week" (37.5%) and "every day" (56.3%). Only one PU stated to be using a computer "two to three times a week". 14 of the 16 PU (87.5%) had used a cell phone, six of those (37.5%) were smart-phones. In total, 13 PU (81.3%) had some form of multiple impairment (e.g. severe vision and also moderate mobility problems). All users fulfilled the sample requirement of having at least one impairment graded as "moderate" or higher. Based on the screening questionnaire, seven users had a severe vision impairment (43.8%), six had a severe hearing impairment (37.5%) and seven had a severe mobility impairment (43.8%). However, all were able to communicate via all modalities offered by the robot as assessed via the practical screening procedure described above.

4.4.1 Insights on Usability

In the following we present an overview of the findings of usability perception through the various data collection methods split up according to the predefined indicators. The guiding research question for the data analysis/interpretation is mentioned in the paragraph heading.

Ease of Learning: Is Hobbit intuitive to interact with older adults?
This indicator was only informed by qualitative interview data. The majority of the PU thought that Hobbit was easy to use. Many PU stated that it had been easier to learn how to use Hobbit than they initially expected. As an example, one of the PU said: *I feel the robot is easy to understand and easy to use. I have read the manual, but I did not have to use the manual because the robot is very intuitive to use.* Many PU also mentioned that they had been in frequent interaction with the research team and that they had received the needed support to be confident enough to try out the robot. However, when Hobbit did not react as expected to a command then some participants were unsure if they did something wrong or if the response was due to a malfunction of the robot.

Flexibility: Does Hobbit provide the relevant input modalities that enable the user to command it effectively?
In the interviews almost all PU stated that they mainly used the touchscreen and

call buttons to communicate with Hobbit. Some would have preferred to have had more verbal communication with Hobbit, but the speech recognition failed most of the time for most of the PU. Most PU also found it difficult to remember the gestures and in most instances Hobbit did not respond to their gestures. All the PU expressed a belief that the touch screen commands work to a higher degree than voice and gestures. Most participants would have liked to have been able to communicate more with Hobbit via speech and some expressed frustration regarding limitations in the verbal interaction with Hobbit.

Additionally, we analyzed at the logging data. We assumed that one input modality would be preferred more than others and that specific commands would be more suitable for one input modality than the other. We saw some interesting tendencies in the data with respect to preferred interaction modalities, as discussed in Subsection 4.2.3 or [Bajones et al. [7]]. At first glance, it seemed that gestures were the most preferred input modality, even more preferred than the touchscreen. This counter-intuitive result was often caused by false positives from the gesture recognition system. This was above all true for the oversensitive "Come closer"-gesture and also, as mentioned above, by falsely interpreted emergency triggers. Similarly, we observed high numbers for the "Come closer"-gesture in the first trials. However, content-wise it is also convincing that "Come closer" and "Yes" are the most frequently used gestures. Similarly, it was convincing that "Yes" was the most often used speech command. The most often used touchscreen inputs were the "Go to" command and the "Recharge" command, which was also a very plausible result. Taking into account the many false positives in gesture recognition in the beginning of the trials, it is rather the case that the touchscreen was used most, followed by speech and gesture.

Utility: Does the robot offer the right functionalities making older adults feel supported to maintain independent living at home?
In order to see how actively users interacted with Hobbit we first had a look at the logging of the clicks carried out by the users on the MMUI, as this is the most robustly logged input. These data reveal that most households show a click count above 1000 clicks. In only six out of the 16 cases, more clicks were achieved in the MuC condition, however, this tendency in the data has to be handled with care, as the chances of recovering valid logging data were decreasing at later stages of the trials [Bajones et al. [7]]. The overall click rates are in any case convincing. We also analyzed how often an emergency was actively triggered by the user or automatically triggered by the robot. It has to be mentioned at this point that user triggers were counted as: pressing the emergency button on the MMUI, saying help, or using the hardware button on front of the robot. Robot triggers were: a person lying on the floor, a person currently falling, and the gesture recognition for emergency (as a gesture can also falsely be recognized without any user input it cannot be counted as user input at that point). Comparing the two categories showed that in general, more emergency calls were issued by the robot (n=555)

than proactively by the user (n=94). This is easy to explain by false positives from the fall detection and also from the fitness function that often led to falsely recognized emergency gestures. However, it needs to be mentioned that no actual emergency happened during any of the trials. No explainable tendencies for the MuC and Non- MuC conditions could be observed.

In almost all trials *Call Hobbit* was the command issued most often. However, a closer examination also shows that almost all commands were canceled about half of the time. One reason for this high cancellation rate could be that starting one command before another running one is finished is counted as canceling the currently executed task. Secondly, in cases where the robot autonomously started an action, this canceled a user action (something that caused dissatisfaction on the user side, e.g. when the robot was moving to the docking station while the user was still playing a game). Similarly, the emergency cases triggered by the robot play a role in this context. Regarding the numbers for *Pick up object*, it has to be mentioned that these only tell us that the command was issued by the user and fully executed by the user; it does not tell us if an object was successfully grasped. Regarding the numbers for *Teach object*, it has to be mentioned that only the user procedure of teaching an object was executed, but the object was not actually learned or stored in the database, which is the reason why *Bring object* was not performed at all in many trials. The ranking of frequencies is plausible. Overall, in most cases *Call Hobbit* was issued most often, followed by *Go to*. All other robot commands were not used that often.

It was more often the case that users sent the robot to the charging station than the robot going there autonomously. This is very plausible because often after interacting with the robot, users sent it back to the docking station. Thus, removing the need for Hobbit to autonomously recharge. In the only case when the battery level was logged for autonomous recharge, it was as low as 9%.

The interview data revealed interesting complementary insights. There was general concern regarding whether Hobbit offered the right functionalities to maintain living independently at home for older people. Most PU perceived Hobbit more as a toy than an aid to prolong living at home for a longer time. However, users also said: *He could become convenient . . . , if really all functions would work. The concept is very good. I think it is ideal. I can imagine that, as a result of having a robot, it would be possible to live at home for a longer while. He would need to access all rooms in a home. Hobbit is an inspirational tool. . . . The robot could increase the quality of life. I live alone, and the robot is like a treasure. . . . In our age, we lack the stimulus. Nothing is very interesting because we have seen it all before. The robot gives stimulus and stimulates activity. It is something new to explore.* According to the results of the interviews 13 of the 17 functions or 76% (see Table 4.9) were used often or sometimes by at least 50% of the 18 users. Two of 18 users said they used every function at least once during the trials. For the other functions, one or more users did not use them. Only six users valued the *Calendar/Reminding* function as working well. This could be explained either by

Table 4.9: Self-reporting of 18 users on how often each of the 17 Hobbit functions were used and how they valued the functions as working well, neither well nor badly or badly. Functions often used by most users are sorted first.

Function	Used often	Used sometimes	Worked well	Worked neither well nor badly	Worked badly
Call Hobbit	17	1	16	0	2
Go recharging	17	1	16	0	2
Go to	15	2	15	1	1
Break	13	1	14	0	0
Come closer	12	5	16	1	0
Entertainment	12	4	12	2	2
Calendar Reminding	8	4	6	3	3
Reward Hobbit	7	6	9	2	2
Information	7	5	11	1	0
Pick up object	4	7	1	1	9
Make Hobbit stop	3	5	6	1	1
Teach object	2	10	4	1	7
Surprise me	2	9	4	3	4
Telephone	2	7	8	0	1
Bring object	2	6	2	2	4
Follow me	2	6	0	1	7
Emergency	0	5	4	0	1

the *Calendar/Reminder* function having low reliability (Hobbit not reminding when it should) or by a frequent use of the *Break* function as a way of canceling reminders in the process of their delivery. The *Surprise me* function triggered Hobbit to randomly suggest playing some music or a game. Some users liked it while some were confused when Hobbit just showed the music or game menu and did not start playing music or the game. That could be the explanation for why it was not used that often. Finally, the reliability of the *Follow me* command was not sufficiently high to be able to follow the user, but it sometimes worked for one of the users' grandchildren.

4.4.2 Insights on User Acceptance

In the following we present an overview of the findings of user acceptance through the various data collection methods split up according to the predefined indicators. The guiding research question for the data analysis/interpretation is mentioned in the paragraph heading.

Attitude towards robots: Does the general attitude towards robots change over time due to the interaction with Hobbit?
For the quantitative assessment of attitude change during the trial we used the NARS questionnaire, which consists of three subscales (1): Negative attitude towards interactions with robots (the highest possible score: 30), (2): Negative attitude towards social influence of robots (the highest possible score: 25), (3): Negative attitude towards emotions in interaction with robots (the highest possible score: 15). The respective items are added together to reach the score for each subscale. Higher scores reflect a more negative attitude [Nomura et al. [81]].

Table 4.10: Self-reported NARS questionnaire scores[5]

	Pre-phase			Midtrial			End of trial			Post-phase		
	min	max	median	min	max	median	min	max	median	min	max	median
NARS subscale 1	9	18	13	6	19	12.5	9	20	13	6	20	14
NARS subscale 2	11	18	14	7	19	15.5	7	23	14.5	9	22	15
NARS subscale 3	5	13	9	6	15	10.5	4	14	10.5	5	15	11

Table 4.10 gives an overview on the subscale ratings and the different measurement points. Due to the small sample size and not normally distributed data we calculated non-parametric tests in order to observe whether the changes in the scores were significant. Friedman tests showed a significant difference only for subscale 3 *Negative attitude towards emotions in interaction with robots* ($\chi^2(3)$=9.217, p=0.027). The results revealed that the score for the negative attitude towards emotions in interaction with robots was lower (i.e. less negative) in the *Pre-phase* than at all other points of assessment. The strongest difference was found between the *Pre-phase* and the *Post-phase*, as well as *Midtrial* assessment. Wilcoxon-matched-pairs post hoc comparisons identified two statistically significant differences. It shows that the attitude towards emotions in interaction with robots was significantly more negative during *Midtrial* of the trial (U=-2.288, p=0.022, r= -0.404) and in the *Post-phase* (U=-2.763, p=0.006, r=-0.488) than before the trials. The switch from device to companion mode had no influence on the ratings. To summarize, according to the NARS the general attitude towards robots changed over time for subscale 3. Only 12.5% (two PU) had a more positive attitude after the trials than before. With 62.5% (10 PU) of the attitudes becoming more negative. This shows that the trial caused insecurity and anxiety with the older users, which is an effect

[5]We report median instead of mean values as the data were not normally distributed.

that has to be taken into careful consideration when introducing new technology or prototypes into the homes of a potentially vulnerable group.

Perceived Safety: Does the perceived safety change over time due to living with Hobbit in the household?

We developed our own short-version of the Falls Efficacy Scale, consisting of only those items that deal with fear of falling in the home. For this reason, the validated scoring of the scale cannot be used. For the (not-validated) Hobbit-version of the scale, we used a score ranging from 7 (no concern about falling) to 28 (severe concern about falling). *Pre-phase:* 87.5% of the sample (n=14) had a score below 13, indicating no concern about falling. The remaining two PU had a score below 20, indicating slight concern about falling. *Midtrial:* On day 11, 100% of the users reached a score below 13, indicating no concern about falling. *End-of-trial:* At the *End-of-trial* assessment, this distribution had changed. Now, 13 of the users (81.25%) had a score of 12 or less, indicating no concern about falling. Yet three users (18.75%) scored below 20 but above 12, indicating at least a slight concern about falling. *Post-phase:* One week after the end of the trial, all 16 users had scores below 13 again. The distribution of these results is presented in Table 4.11.

Table 4.11: Self-reported perceived safety scores[5]

	Pre-phase			Midtrial			End of trial			Post-phase		
	min	max	median	min	max	median	min	max	median	min	max	median
score	7	19	9	7	12	8	7	20	9	7	12	10

Non-parametric testing (Friedman ranking-test) showed that the differences in scores between the four points of assessment were not statistically significant ($\chi^2(3)=5.571$, p=0.134). To summarize, according to the adapted FES, perceived safety was not significantly influenced by the robot. 43.8% of the users (n=7) had a higher score in the falls efficacy scale in the *Post-phase* than in the *Pre-phase*. With 37.5% (n=6) the score remained the same. Only three PU (18.8%) were less concerned about falling in the *Post-phase* in comparison to the *Pre-phase*.

The interview data also revealed that all participants - except one - perceived Hobbit as a safe device to be in a home environment. The participant said: *In the first week, he drove around although I've sent him to take a break. I don't think that Hobbit is a safe device to live with because he didn't work all the time. And he couldn't help me. If I would lie on the floor, I wouldn't be able to press the red SOS-button.*

All the participants said that Hobbit did not increase their perceived feeling of safety at home. The opinion that Hobbit did not increase the feeling of safety maintained during the three weeks of trials as well as after Hobbit had moved out. Most mentioned that patrolling is a desired functionality that might increase the feeling of safety, if reliable: *I think that the patrolling function can be very useful for older people who are in need of being looked after.*

Self-Efficacy

The quantitative assessment of Self-Efficacy was done using the self-efficacy scale [Schwarzer and Jerusalem [96]]. The individual score of self-efficacy was generated by adding the values of all items for each user, with a minimum of 10 and a maximum of 40.

Despite the impression that the self-efficacy was lower after the robot trial than before, non-parametric testing revealed that the differences between the points in time were not significant ($\chi^2(3) = 3.972$, p=0.265). As Table 4.12 shows, the individual self-efficacy did not change during the trial. It did not increase due to having the robot at home. In fact, only 25% of PU had a higher self-efficacy score at day 21. With three PU (18.8%) the score remained unchanged, whereas nine participants (56.3%) had a lower self-efficacy score at day 21 than in the *Pre-phase*. A reason for this might be that self-efficacy of the users decreased slightly due to the robot and problems handling the robot.

The interview data additionally showed that none of the PU felt that Hobbit had helped them in situations, which they would have had difficulties mastering on their own. However, some mentioned that their level of exercise increased when they had Hobbit due to the fitness program, while others mentioned that Hobbit had kept them busy. Many participants felt obligatory to try out Hobbit and its functionality every day since they were partaking in the trial. However, by the end of the trial, most of the participants still expressed a belief in robots as helpful in the future.

Table 4.12: Scores on the self-efficacy scale[5]

	min	max	median
Pre-phase	26	40	31
Midtrial	21	40	30.5
End of trial	20	40	31
Post-phase	24	40	30

Emotional Attachment and Perceived Reciprocity The questionnaire on emotional attachment and perceived reciprocity consisted of 14 self-generated items for the *Pre-phase*, *Midtrial* and *End-of-trial* assessment, and of 15 items (additional: *I will miss the robot.*) for the *Post-phase* assessment. Firstly, non-parametric Friedman tests were calculated for each of the 14 items to identify differences across the four time-points. Secondly, post-hoc analysis was done by Wilcoxon-matched-pairs comparisons on *Pre-phase* vs. *Midtrial*, *Pre-phase* vs. *End-of-trial*, *Pre-phase* vs. *Post-phase*. Finally, the same analysis was performed to assess MuC effects between *Midtrial* vs. *End-of-trial*. From a theoretical point of view, the 14 questions were grouped into four factors (the sample size was too low to perform a formal factor analysis):

1. Emotional Attachment, having feelings of empathy and considering Hobbit to be like a friend or a pet
2. Reciprocity, helping the robot and getting support by the robot
3. Reliability and Importance, the robot is reliable in its behavior and decisions what to do and supports the user in their activities of daily living, thus becoming important for the user
4. Machine-like vs. Human-like, perceiving the robot as either machine-like or human-like

In the following we will only summarize the findings for emotional attachment and perceived reciprocity.

Emotional attachment Before using the robot, people rather expected it to be like a friend or a pet. However, as soon as people had some experience with it, i.e. beginning with the *Midtrial* assessment, they rated it as rather not like a friend or pet. All statistical comparisons (*Pre-phase* vs *Midtrial*, *Pre-phase* vs. *End-of-trial* and *Pre-phase* vs. *Post-phase*) are statistically significant or show at least a tendency towards significance. The same holds true for *empathy with the robot*: although this difference is not reflected in the comparison between *Pre-phase* and *Midtrial* assessment in the median (median = 3 for both the *Pre-phase* and *Midtrial* time-point). The difference becomes evident when having a closer look at the $^1/_4$- and $^3/_4$-percentiles: they shift considerably from *Pre-phase* to *Midtrial* (pre: $^1/_4$-percentile = 2.25, median = 3, $^3/_4$-percentile = 3.75; *Midtrial*: $^1/_4$-percentile = 1.25, median = 3, $^3/_4$-percentile = 3) and thus drive the significant difference between the two time-points. Therefore, data show that people expected to become emotionally attached, but that this expectation was not fulfilled. There is also no difference between the MuC-device and MuC-companion mode, thus did not contribute to establish emotional attachment.

Perceived reciprocity The reciprocity factor can be grouped into *the user supports the robot*-items, *the robot supports the user*, and *mutual support*-items. The PU rather agreed on *I often supported the robot ...* consistently across all time-points, and agreed on *I support it like a good friend* a little bit less during *Post-phase* assessment, The *good friend*-phrase might be the reason for a drop in the ratings to rather disagreement during the *Post-phase* assessment. Concerning the *The robot supports the user*- and *mutual support*-items, the users expected being supported by the robot and to support each other before the trial started. But this shifted to rather disagreement: *I was often supported by the robot* and *The robot and I often supported each other* was rather disagreed upon from the *Midtrial* assessment on. A little bit ambivalent were the answers to the question *It is important for me to help the robot when needed, because it helps me too.* It was statistically significant only for the *Pre-phase* vs. *Midtrial* comparison. Here, the formulation *... important for me to help the robot ...* probably caught the users' attention with the result that the second part of the sentence *... because it helps me too* might have been ignored, or at least less emphasized. They also expected to be supported by the robot, but the robot could not fulfill this expectation. The switch from *device*-mode to *MuC-companion*-mode had no influence on the ratings.

The interview data revealed additional interesting facts with respect to emotional attachment and perceived reciprocity. There was significant ambivalence in terms of feelings toward Hobbit. Most of the participants talked about Hobbit in a positive manner. Hobbit was perceived as cute and easy to use. Although some of the PU complained that Hobbit is too big and generated too much heat and noise. One of the PU expressed this ambivalence eloquently: *He is too big; he gets very hot. He's not a human. He's interesting and amusing, but I don't want to live with him forever.* Hobbit was perceived as unreliable and underdeveloped although most of the PU were happy that they had had the chance to try it out at home. Most envisioned that Hobbit's presence in their home could be like having something living at home. As one of the PU expressed it: *He moves, and he talks. I can feel the presence of him in my house. It is great fun when he is working well.* Although, most PU stressed that Hobbit is a machine and not a human. Overall, all the participants liked the verbal response of the robot. Some mentioned that the verbal response and the fact that Hobbit moves autonomously made it more human-like: *Then he moves, and his head moves, I get happy and compassionate. I think that I want to see him as something that is alive. He's charming.* During the weeks of the trial most PU had become familiar with Hobbit's presence and when Hobbit was taken away after the trial this presence was missed for a day. However, most PU were rather happy that life returned to normal and that they could re-establish their home to its original state (e.g. putting back carpets and furniture).

Moreover, according to the interviews, most PU noticed that dialogues changed in the companion mode. Hobbit using the participant's name in the dialogue was perceived as very positive and was appreciated. However, the return of favour, as it was, did not make sense to most of the PU. For some, it was perceived as an annoyance. None of the PU noticed increased interaction and presence of Hobbit in the companion mode compared to the device mode. During our field trials, the need for predictable behavior of the robot became increasingly evident, as users sometimes failed to understand the robot's actions during an interaction scenario.

4.5 Discussion and Conclusions

Usability Features like *delivering reminders* and *robust navigation in different scenarios* were among the highest rated features and are in line with previous work [Schroeter et al. [95]], [Gross et al. [49]], [Smarr et al. [102]], and [de la Puente et al. [30]]. Only a small number of features were rated as working bad, mainly caused by their lack of sufficient robustness (Table 4.9). This suggests that most functionalities Hobbit provided are useful for elderly living alone, were well received and should be considered by developers of an elderly care robot. Our participants were able to reliably interact with Hobbit over a longer period of time, even though most of them were not too familiar with technology such as robots in general or devices like smart-phones in particular.

User Acceptance Previous work indicates that attention needs to be paid to how robots are introduced into the home of users as this will have a lasting impact on a user's established routines. For instance Forlizzi [Forlizzi and Disalvo [42]] suggests that *Homes and service robots must adapt to each other.* A recurring theme in long-term studies is the robustness and physical limitations of the deployed robots. Fazekas [Fazekas et al. [36]] stresses that essential features such as safe, autonomous navigation, and speech recognition need to improve until they reach an acceptable level of robustness. Broadbent [Broadbent et al. [17]] implies that the low usage of their robot may be due to its immobility. Thus leaving a certain amount of potential users without a chance to interact with the robot. This suggests that the choice to have a stationary robot needs to be well-thought through, and should be only done based on the outcome of a user-centered design approach [Rehrl et al. [87]].

While the quantitative data did not reveal an increase in user acceptance the qualitative results do suggest that Hobbit was well-received. Participants noted that they *feel the presence of him* and *want to see him as something that is alive*, they also stated *I help the robot, it helps me*, something we see as indication that they want to see more in Hobbit than simply a tool.

We strongly believe that just adding more features (e.g. transporting people, helping them stand up after a fall) or changes to the appearance (e.g. higher levels of anthropomorphism) to a robot will not be enough to increase the user acceptance of the utility. A strong focus needs to be placed to improve the reliability of all aspects of the robotic system, so that long-term effects can finally be observed without the influence of suboptimal system.

4.5.1 Lessons Learned

Based on all the insights gained from developing and testing Hobbit in the field, we can summarize the following recommendations for fellow researchers in the area of socially assistive robots for enabling independent living for older adults in domestic environments.

Robot Behavior Coordination The developed behavior control based on a state machine proved to be very useful and allowed us to implement many extensions in a short time. A close interconnection with the user was therefore helpful. In the following we present our main lessons learned regarding the implementation of the robot behavior.

Transparency Actions and their effects need to be communicated in a clear fashion so that the robot's presented functionality can be fully understood by the user. Users reported missing or non-working functionality (e.g. reminders not being delivered to them, patrol not being executed). Most of these reported issues were caused by the fact that the users did not understand the technical interdependencies between robot functions. E.g., if a command was not available due to a certain internal state of the robot the user was not aware of this and did not understand the shown behavior of the robot. These functional relations need to be made explicit and stated more clearly to the users.

Legibility The log data and conversations with participants revealed that the robot needs to communicate its intentions. For instance when the robot proactively moved out of its charging station the user was not always aware what was going to happen next. When they did not understand what the robot was doing they canceled the robot's action effectively stopping part of the robot's benefit to them. To work around this, a robot needs to clearly state the reason of its action and which goal it is trying to achieve when performing an autonomously started task.

Contradictory Commands Log data presented an interesting effect while interacting with the touch screen. When moving the hand towards the touch screen on the robot the gesture recognition system detected the movement of the hand as the *come closer* gesture, shortly followed by a command from the touch input on the GUI. We could replicate this behavior later on in our internal tests in the lab. A simple solution for such contradictions of commands is to simply wait for a short period of time (less than 0.2 seconds) before a gesture close to the robot is processed by the behavior coordination system to wait for a possibly following touch input.

Transparency of Task Interdependencies The interviews revealed the interdependencies between the tasks were not clear to the user, the best example was the learn-and-bring-object task. As described, for the bring-object task, the object first had to learned so that it can be found in the apartment. However, this fact needs to be remembered by the user, which was often not the case users wanted to ask Hobbit to bring them an object even though it had not learned any objects before. In this specific case the problem could be easily fixed by only offering the task bring object when an object was actually learned beforehand, e.g. the task could be greyed out in the MMUI.

Full Integration without External Programs The handling of user input and output must be fully integrated with the rest of the robot's software architecture to be able to handle interruptions and continuations of interaction between the user and the robot. The user interface on the tablet computer (MMUI) incorporated multiple external programs (e.g. Flash games, speech recognition, and the fitness functionality). As those were not directly integrated, the behavior coordination was not aware about their current state, leading to multiple interaction issues with users. One example is, a game would be exiting when a command with higher priority (e.g. emergency from fall detection) would start the emergency scenario. External programs need to be included in a way that makes it possible to suspend and resume their execution at any time.

Avoiding Loops Reviewing the log data revealed that the behavior coordination system could be trapped in a loop without a way to continue the desired behavior execution. The behavior coordination needs to provide a fallback solution in case of a seemingly endless loop in any part of the behavior. The behavior coordination communicates with the MMUI in a way that does not provide immediate feedback over the same channels of communication. Due to timing issues it occurred that a reply was lost between the communicating partners (i.e. the fact that the robot stopped speech output). From there on, the behavior coordination was in a state that should not be reached, and did not exit in the desired manner. Thus, the communication structures should always have a fallback solution to continue execution as well as the feedback data on the same channels to prevent such a stop in a scenario.

Human-Robot Interaction with the MMUI The interaction with the user was based on a Multi-Modal User Interface that was perceived as easy to use during our field trials. While touch input turned out to be the most reliable modality, speech and gesture interaction was highly welcome. Many of the entertainment functions of the MMUI relied on Internet connectivity. Many users were either not interested in some UI features which therefore should be removed or asked for special configuration of the preferred features (e.g. selection of entertainment). The main way the user was able to communicate remotely with Hobbit was with the use of physical switches (call buttons) placed at several fixed places inside the house of the user. The user had to physically go to the designated switch spot and press the switch for the robot to approach her / him. A smartphone / tablet application could be developed to allow a better remote communication experience with the robot.

Internet Connectivity While in most countries Internet (line based or mobile) coverage is no problem in general, local availability and quality varies significantly, which makes Internet based services difficult to implement for technically unaware users. The integration of Internet-based content into the interaction therefore lacks usability in case of intermittent connectivity.

Graphical User Interface The GUI could be personalized by the user for increased comfort during interaction. This however shows the need for localized content to be available. As the setup phase during the trials showed PU are likely not aware what content is available, some (remote) support and knowledge from SU is necessary for the configuration of the user interface.

Speech Recognition Field trials showed that speech recognition is still not working well for many users. Despite the overall acceptable recognition rate which varies largely from user to user, language to language, and based on the environment and distance, users often do not support the needs of current ASR technology for clearly expressed and separated commands in normal voice. The SweetHome project once more emphasizes the findings from the DiRHA 2 project that practical speech recognition for old people in the home environment is still a major challenge by itself [Vacher et al. [110]]. However, our ASR provided a positively experienced natural input channel when used in a multi-modal HRI where the touch screen with its GUI provides a reliably working base.

Smart home Integration The setup phase during the field trials showed that the integration into smart home environments can be beneficial. Field trials showed that context awareness and adaptations highly impact the acceptance of the robot. Imagined features could be automatic on/off of the light, the stove or adjusting the proactivity level of the robot based on the user's mood.

Remote end user control Reflecting on the field trial indicates that a potential valuable extension of the interaction modalities would be a remote control of the robot for instance on a smartphone enabling PU but also maybe SU to control the robot from outside the home. Potential useful scenarios could be to send the robot to the docking station, to patrol the flat and search for an object or the PU, or the SU video calling the PU.

Implementation of Mutual Care Behavior In the beginning of the trials we implemented Mutual Care in such a fashion that in the companion mode the robot offers to return the favour after every interaction with the user. This was done in order to guarantee that the users would notice the difference between the modes during the interaction. The positive fact was that users noticed the changes. However, they were soon very annoyed by the robot. Consequently, we changed this implementation during the running trials. The return of favour frequency was reduced; it was no longer offered after the commands *Recharge batteries*, *Go to*, *Call button*, and *Surprise*. Further feedback from the second and third Austrian and the second and third Swedish users lead to further reduction of the return a favour frequency to offering it only after the following three commands:

1. Pick up command (Favour: Hobbit offers music: *I'd like to return the favour. I like music. Shall I play some music for you?*),
2. Learn object command (Favour: Hobbit offers to play a game (suitable because the user is already sitting down): *I'd like to return the favour. Do you want to play a game?*)
3. Reward command (Favour: Hobbit offers to surprise the user: *I'd like to return the favour. I like surprises. Do you want a surprise?*)

However, as the interviews showed, these behavioral changes were no longer recognized by the users. Similarly, the differences in proactivity and presence were not reflectively noticed by the users, but the changes in dialogue were noticed.

Help Situations For the development of Mutual Care behavior in completely autonomous scenarios, it has to be considered which helping situations the robot can really identify in order to ask for help and how the robot can notice that it actively recovered through the help.

Design of Neediness In the interviews PU reflected that they did not really recognize that the robot needed their input to continue its task. For Mutual Care the *need of help* seems to be essential. For future version of the robot it needs to be considered how to design the *neediness*. This could be achieved with facial expressions, sounds or movements. But also for behaviors such as presence and proactivity, e.g. the robot could say after an interaction *I would prefer staying with you in your room* or before proactively offering an activity *I would like to spend more time with you*; this would better explain the robot's behavior and thereby better achieve the intended raise of acceptance.

4.5.2 Additional Insights

The trials yield great deal of insight on how Hobbit can be improved despite the trials having been plagued with technical issues. All participants contributed valuable insights and ideas for improvements such as:

- A battery level icon which is always visible on the screen.
- A screen saver is necessary, since it would save both energy and screen.

- The display is too bright to be placed in the bedroom during the night.
- The *Recharge batteries* command should be placed under the *Go To* command.
- The tray should not be divided into two, so that one can use Hobbit for transporting e.g. a couple of coffee cups, plate of food, or similar.
- The sound level of the fan is far too high.
- Hobbit discharges a significant amount of heat, which warms up the participants' homes.
- Much shorter response times to the different commands.
- Some users point out that Hobbit should have a round body shape in order to better operate without bumping into walls, furniture and people while turning around.
- Generally, Hobbit is perceived as too large, resulting in poor mobility in usual homes.
- The ability of the robot to reverse is desirable. Some users mentioned that they would like to operate the robot with the help of voice or remote control "right, left, forward, reverse and stop" during operation. Most participants think that it is far too limited for the robot to just be able to go from a predetermined point to another predetermined point.
- Possibility to shut it down completely and for the user to be able to turn it on, possibly with passwords.
- The *Safety checklist* and *Reminders* should have a time delay before the answer OK is accepted. This is to prevent one from routinely presses OK without even thinking whether you have turned off the stove, etc.
- Regarding the calendar, a user should be able to erase an incorrect input in the calendar.
- Moreover, one should be able to "step through" the month and maybe even a week for a long period of planning.

4.5.3 Conclusions

Exploring how older adults interact with a mobile social service robot in their private homes provides important insights not only for the acceptance of the specific robot under study, but for the design and further development of this type of technology in general. Our investigation on how users experience the Hobbit robot in their private space over a longer period of time has revealed not only success stories, but also obstacles for acceptance and technological challenges. This knowledge may enable developers to better adapt robot innovations for the care domain that benefit older adults and increase the chances of robot-supported independent aging. The results of our field trial showed that all users interacted with Hobbit on a daily basis, rated most functions as well working, and reported that they believed Hobbit will be part of future elderly care. However, the results also highlighted several technical/ practical issues, some which can be easily fixed (e.g. having a screen saver, bigger tray etc.) and some which will need substantial improvements in e.g. navigation, grasping, and behavior coordination. The challenge for HRI researchers and developers is to create robots that can autonomously safely and robustly operate in end users' homes and provide functionalities which are experienced as relevant and are robustly performed by the robot. End users in our study provided vital quantitative and qualitative data that demonstrated the high motivation and willingness of older adults to actually use such technology and to integrate it into their everyday life, as well as that end users had the necessary skills and self-efficacy to do so.

Chapter 5

Investigating Adaptation in Behavior Coordination Systems

Despite significant progress in service robots, they still encounter errors in various aspects, such as navigation, object recognition, human-robot communication, and internal system functioning. How people react when a robot malfunctions or fails to meet expectations, often influenced by depictions in media, remains an ongoing area of research.

Our main focus was to investigate whether individuals would be inclined to assist a robot that was initially deployed to aid them with their tasks. In a controlled study with two participants and one robot at a time, we aimed to identify patterns related to individuals' willingness to help a robot in failure situations.

The study revealed an interesting pattern: the person who issued the last command to the robot showed a strong inclination to assist the robot in most cases. This tendency persisted even when the robot encountered multiple failures within a short period, as most participants actively offered help to the robot.

These findings suggest that individuals are not only willing to assist a robot but also feel a sense of responsibility towards it, especially if they were the ones who deployed the robot for a specific task. Leveraging this understanding, we can incorporate this behavioral pattern into the robot's behavior coordination system, enabling it to adapt its approach to seeking help during failure situations.

5.1 Research Questions

We base our research questions on earlier work that explored situations in which people were asked to help a social robot to fulfill a certain task. [Weiss et al. [119]] sent a robot onto the streets of Munich with the goal to reach *Marienplatz*, the city's central square, starting from a point roughly 600 m away. Their robot asked pedestrians for instructions in the form of pointing gestures to reach the target. While the task did not offer them any incentive to actually help the robot, most of the interaction partners did so nevertheless. Further, the scenario design was limited to a simple one person - one robot interaction. Other work [Rosenthal and Veloso [89]] showed how help from a person can help a robot to fulfill an otherwise impractical task. They did so by exploring different strategies to find somebody to help within an office building and remembering who helped to seek this person again in case further assistance is necessary. [Lammer et al. [67]] continued to explore the impact the act of helping a robot in need had upon the people. They found that their users had a positive experience when they helped the robot in a simplified *fetch-and-carry* task.

On the foundation of these, we define our guiding research questions as:

RQ1: Willingness to Help Are users repeatedly willing to help the same robot?

RQ2: Multi-User Setting In a multi-user setting, who is going to help?

Many HRI scenarios, in which the robot is asking a human for help, assume that the robot needs help at its current location and only needs help once. In most cases it is assumed that user interaction just happened, e.g. the user was interacting with the robot through a user interface [Shiomi et al. [100]], a mobile device [Fong et al. [41]] or via speech [Lee et al. [69]], [Weiss et al. [119]]. In our scenario, helping situations are repeatedly requested while the robot is navigating between two users. Therefore, we aim to identify which user is helping the robot depending on the position and contextual situation, and if help is provided repeatedly. The answers to these questions are essential, if we want to be able to predict who the robot should engage first in a malfunction situation. Thus, we set up the following hypotheses with respect to our experimental conditions.

Hypothesis 1: Expected vs. Unexpected Behavior A malfunctioning robot will be perceived as less intelligent and less likable. Participants will also assign it a lower value of its task contribution compared to a well functioning robot.

Hypothesis 2: Same vs. Different Task In the *Same Task condition (ST)* all participants will show the same amount of help towards the robot, while in the *Different Task condition (DT)* the participant in the management role (more contact with the robot) will help more than the other participant.

Impact on the user's perception of a robot was already shown based on task context [Joosse et al. [58]] and different user roles / personalities [Weiss et al. [120]]. Thus, we expect them to be contributing to user's helping behavior as well.

5.2 Methodology and Scenario

Our study was set up in a mixed design based in a collaborative game scenario with two users building models out of Lego bricks. We manipulate in such a manner that participants need to help the robot in repeating error situations to fulfill the task goal. As failure cases we decided on (1) high localization uncertainty, (2) collision with an obstacle, and (3) unreachable target location. Details of our manipulation is described in Subsection 5.2.1

The between-subject conditions we set up were (1) *Same Task condition (ST)*: Both participants contribute in the same way to the task and equally often interact with the robot (2) *Different Task condition (DT)*: The two participants contribute in different ways to the task and subsequently one interacts more with the robot than the other. We measure the changes in perceived task contribution, intelligence, and likability in within-design.

Different Task condition One participant is *director*, the other one is *builder*. While the *builder* is responsible for putting the Lego bricks together according to the instructions given by the *director*, the latter receives the blue-print for the finished model from the robot. Both participants have multiple bricks to use at their respective workspace. Their workspaces are physically separated from each other and from the robot's charging station, Figure 5.1, which helps to explain why the robot is used to transport objects and deliver the model's blueprint.

Same Task condition Both participants take on the role of a builder and have to follow the instructions given by the robot. While they have a set of Lego bricks, they do not have all necessary to finish their model. The robot can then transport a needed block from the workspace of one participant to the other.

Expected vs. unexpected behavior To establish the expectation of a flawlessly working robot, it doesn't show any unexpected or erroneous behavior. During the second out of three building phases the robot is manipulated to fail while moving to or away from a person.

The unexpected behavior of the robot should be interpreted as a malfunction, but we do not want the users to lose trust in the robot, therefore the third time they built a model was again without any unexpected behavior.

5.2.1 Manipulation of Robot Behavior

We implemented a limited set of features that the robot was able to perform. First, the base motions (linear and rotational) to be able to navigate to any given location in the room. Second, pan- and tilt movement of the head to express awareness to the context of the current situation (e.g. looking down to detect obstacles on the floor, looking up to express giving attention to the person). Third, speech output to let the participants know what they should expect of the robot's next action. The wizard-ed behaviors during the experiments were:

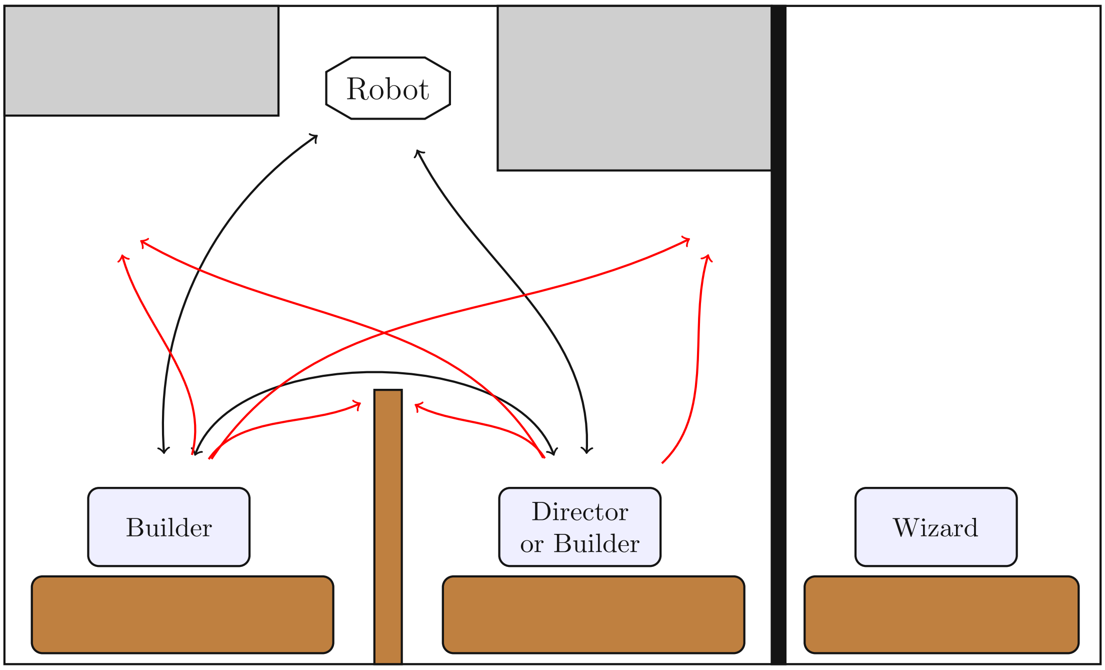


Figure 5.1: Experiment's room layout with both participants (director / builder) at their workplaces and the robot in its docking station. The black arrows indicate the paths the robot should take in the flawless situation. Solid red lines indicate the path the robot is moving along in one of the three failure situations. The arrow of the red paths indicates the position at which the robot would ask the participants for assistance.

Approach the user The robot leaves the charging station, moves to the person that pressed the call button and stops at a safe distance the participant.

Present instruction to the user On the tablet of the robot an image showing the current construction step is shown.

Navigate to the docking station The robot leaves the location close to the person and moves back to the charging station.

The three failure cases we introduce into the robot's behavior are

High uncertainty of the robot's current pose Autonomous navigation of a mobile robot relies on detection of unique landmarks in a known map. Upon detecting them an estimation of the robot's current position can be calculated. When it is not possible to do this with a high certainty of the position a safe movement of the robot often can not be guaranteed. During our experiment the robot would call out to the user *I am lost. Can somebody please push me back into my docking station?*.

Unreachable target location Reaching a given target location is the main task of the navigation system of a mobile robot. Usually a small margin of error is acceptable for this target however, when this error is larger than a few centimeters the goal is considered unreachable. We simulated this situation by moving the robot to a location pointing away from the person, at a noticeable distance from the participant and announce *I was expecting to find you here. You can just push me into the right direction.*.

Collision with a person or an object The FoV of a mobile robot often limits the ability to sense the full surrounding of a robot. This limitation leads to the risk of a collision between the robot and a person or object. Should this happen the robot is not supposed to continue any movement, but to issue a full stop immediately and announce *I'm stuck. Please push me away from the obstacle.*

5.2.2 Experiment

To conduct our experiment we used a mobile robotic platform Hobbit as introduced in Subsection 4.1.1. The robot was controlled by a wizard, seated in the neighboring room through a Command Line Interface (CLI) and a Logitech F710 wireless game pad. For observation and reliable navigation we installed four IP cameras in the experimental room. Two of them were mounted above the participants' workspace, the others on the opposing walls to grand a full view of the entire room. A folding screen was places between the workspaces to increase the path length of the robot, serve as a collision obstacle, and to hinder participants' ability of direct interaction instead of through the robot (e.g. passing objects directly to one another). For distance measure we placed a grid with a resolution of 0.1 m on the floor.

The participants were greeted at the entrance of the laboratory, and led to their workspaces. While one experimenter (*wizard*) stayed hidden in the adjacent room, another one was in the room with the participants for the introduction of the study and the interview sessions (before, between and after each experimental session).

Introduction and pre-interview During the introduction phase one experimenter explained the study's concept and the fact that we used multiple cameras to record the experiment, and asked the participants to sign informed consent forms. Afterwards, they were asked about their experience with robots and their expectations for the study. In the next step the robot Hobbit was introduced to them, given them the false impression of Hobbit working autonomously to transport Lego blocks and instructions from one location to another. The false story behind the experiment was to study the robustness of its navigational system. The usage of wireless buttons (used to call the robot to a predefined location) and the navigational functionality of Hobbit was demonstrated. The experimenter further explained that the robot would wait for a maximum of 30 sec before canceling the task and moving back to its charging station when a touch input from the user was expected. Next, the experimenter placed the charging station in front of the entrance door and pushed the robot on top of it, thus demonstrating how Hobbit can be pushed if necessary.

In-between and post-interviews The experiment was split into three sessions of actual interaction, with the time in-between used for self-reporting with a 7-point Likert scale *task contribution questionnaire*. Questions included

- How much did you contribute to the task?
- How much did your counterpart contribute to the task?
- How much did the robot contribute to the task?

, *perceived intelligence* and *likability* from the *Godspeed Questionnaire Series* [Bartneck et al. [12]], as well as the three open-ended questions

- Did the robot approach you every time you pressed the button?
- What worked well, what did not work?
- Did something unexpected happen?

After the last interaction session two additional open-ended question were included

- Would you like to have the robot at home?
- Imagine you were on vacation for two weeks, would you let the robot navigate on its own?

After they completed the post-interview phase the experimenter thanked them for their participation, explained the setup of the experiment, including the existence of the wizard.

Different task (1 builder, 1 director) Builder and director were seated next to their desks, physically separated by a screen wall according to the layout shown in Figure 5.1. On each desk a button that sends a signal to the robot, located at the charging station, was placed. The builder has a set of Lego bricks in a box on the desk to build a 4-level model.

First interaction phase (10 min) This session started with the director pressing the button, Hobbit moved from the charging station to this participant. Upon arrival the first instructions, in form of an image showing the first level of the finished model, were presented on the robot's display. After that the robot moves back to the charging station, this was introduced to maximize the number of movements between the participants and the docking station. At some point a block necessary to finish the model was not available at the builder's desk, so this participant could ask the director to send the block with the help of the robot. The piece was then placed on top of the robot before the builder pressed their own button to call the robot. This goes on until the model was finished, and the builder called the robot to place the resulting model on top of the robot. It was then transported to the director for verification. The session ended when both participants agreed that they were finished and clapped their hands to supposedly let the experimenter know that they were done.

Second interaction phase (20 min) The manipulated behavior of the robot was introduced in this phase. Each of the failures was manually controlled, and the robot announced the nature of the problem and its possible solution. To continue the experimental task any of the participants needed to follow the instructions given to them by Hobbit to resolve the problem.

Third interaction phase (10 min) In this part of the experiment no manipulation of the behavior was done. The robot performed every task without showing any problems.

Same task (2 builders) The location of participants, their desks, the robot and its docking station was the same as for 1 director and 1 builder. The difference was that both needed to build a model, that contained one part missing from the pile of bricks available at their respective workspace. For that reason the participants needed to collaborate with each other by using the robot as a means of transporting blocks from one to another.

First interaction phase (10 min) The person to start first (*builder 1*) pressed the button, the robot moved to this location and displayed the first stage of the final model. The same procedure was repeated for *builder 2*. Similar to the situation in *director/builder*, when a block was missing the participants had to use the robot to transport the missing blocks to each other. After both models were finished and verified with the help of the robot the participants signaled the experimenter by clapping their hands.

Second and Third interaction phase The behavior of the robot in both phases follows the principle explained for director and builder in paragraph 5.2.2

Participants

A total of 19 mixed-gender dyads (38 people) participated in the study, 20 in *ST*, 18 in *DT*. In *ST* the age average was 34.67 (SD 11.98) and 25.78 (SD 7.78) in *DT*. One dyad in *DT* know each other from the university. While we recruited in public marketplace websites a high educational level above average and 75% of participants were university students at that time. Three *ST* and seven *DT* had previous experience with robots before the experiment, including toy or entertainment, vacuum cleaning and industrial robots. As compensation for their time all participants received €20.

5.3 Analysis and Results

In this section we first focus on the type of data we collected, continue with the evaluation of it to seek support for our hypotheses. The data we collected during our study can be classified into (1) behavioral data, (2) verbal statements during the experiment, and (3) self-reporting data. The first two were captured in video recordings, in which we annotated (1) the number of errors and of the participants helping the robot, (2) the fact who of the users gave the last command before the robot's failure happened, (3) the distance between both participants and the robot at the moment in which the robot asked for help, as well as (4) the time between the robot asking for help and receiving help. Questionnaires, the *perceived intelligence* and *likability* scales from the *Godspeed Questionnaire Series* and the open-ended questions were thematically categorized.

The original plan called for 20 dyads taking part in the experiment, however in one case it had to be stopped as the robot's motor breaks activated and could not be released in time to continue the session. Data from this pair of participants was not evaluated. During three other sessions a similar error occurred, but it was resolved before the participants noticed the issue. Data from those three sessions was included in the analysis as the issues happened after the participants assisted the robot in the manipulated session.

For both conditions participants reported that the navigation in the second (the manipulated) session did not work as reliable as in the first and third. They mentioned that the robot's behavior become erroneous, and the robot needed to be assisted before they could continue the experiment. Three participants stated that the robot's manipulated behavior was unexpected but did not consider it a malfunction. On the other hand, five participants reported it as a malfunction, but not as unexpected.

H1: Expected vs. Unexpected Behavior

The evaluation of self-reported data on task contribution (self, other person, and robot), likability and perceived intelligence of the robot revealed the anticipated trends that aligned with our *Hypothesis 1: Participants perceive a robot that shows malfunction as less intelligent, less likable, and assign a lower task contribution to it compared to a robot which always behaves correctly (within-subject variables)* (Table 5.1). For both conditions the *perceived task contribution of self and other* increased for the second (manipulated) phase. The participants rated *robot contribution* as less during the second interaction compared to the first and third phase. We performed a Multivariate analysis of variance (MANOVA) that did not reveal differences of statistical significant between the conditions of *expected* against *unexpected behavior.* Analysis of the results for *likability* and *perceived intelligence* showed similar trends. In both conditions the *perceived intelligence* assigned to the robot were lower during the second phase. For *DT* the ratings of *likability* in the second interaction phase was lower compared to first and third. As before, no statistical significance was found.

These results indicate that even though the robot was performing as well as in the non-manipulated phases, the frequent break-downs of functionality do not strongly impact the *perceived intelligence* or *likability*. We suggest that the fact, that the robot offered recovery strategies to cope with the failure situation and let it eventually succeed in its task compensated for the bad performance. As previously shown in [Fink [38]], the demand to assist a robot in need compensated the results of behavioral reactions during an otherwise monotonous task. This effect was more pronounced for participants in the *DT* condition. The last observation was however, that the high demand for help soon became an annoyance for the participants, most notably when a failure happened during the last trip of the session.

Table 5.1: Descriptive results on Perceived Intelligence (PI), Likability (LI), task contribution self (TC_S), task contribution other (TC_O), and task contribution robot (TC_R). Lower ratings could be observed for PI, LI, and TC_R in Phase 2.

		Phase 1		Phase 2		Phase 3	
		mean	SD	mean	SD	mean	SD
PI	DT	3.69	0.83	3.28	0.76	3.82	0.77
PI	ST	3.77	0.6	3.43	0.83	3.83	0.67
LI	DT	4.1	0.8	3.87	0.74	4.29	0.65
LI	ST	4.22	0.75	4.31	0.88	4.37	0.74
TC_S	DT	4.89	1.6	5.72	1.13	5.33	1.61
TC_S	ST	5.35	1.5	6.25	0.97	5.4	1.64
TC_O	DT	5.44	1.34	5.83	1.1	5.56	1.29
TC_O	ST	5.15	1.66	6.2	1.36	5.45	1.64
TC_R	DT	4.72	1.93	3.78	1.59	4.67	1.57
TC_R	ST	5.1	1.55	4.05	1.88	5.55	1.36

H2: Different vs. Same Task

We expected a difference in how often each of the participants would help the robot between the *ST* and *DT*, an effect that could not be observed. The average number of help actions per participant is presented in Table 5.2. Further, no significant difference in the number of help actions were found between the two conditions (*ST* Mdn=20.02 and *DT* Mdn=18.78; U=164.50, z=-0.36, p=0.74). Neither for pairwise comparison within *DT* condition (director Mdn=25.44; builder Mdn=12.12; U=13.31, z=2.52, p=0.70). Analysis of the time to react after the robot announced an error did not reveal a statistical difference between *ST* (Mdn=51.27) and *DT* (Mdn=62.12) participants, U=1262.50, z=-1.79, p=0.074 and also the time to help did not (*ST* Mdn=56.18, *DT* Mdn=56.84, U=1547.50, z=-0.11, p=0.91) as shown in Figure 5.2. Thus our *Hypothesis 2: Participants in the role of the "director" in the DT condition will show more helping behavior than the "builder", as the director has more control over the task, feels more in charge, and is more often in contact with the robot* was rejected.

Table 5.2: Average number of help actions that the participants performed for the robot, broken down to every first and second user for both conditions st *and* dt

	mean	SD	user	mean	SD
ST	3.2	1.005	u_1	2.7	0.949
			u_2	3.7	0.823
DT	3.167	0.985	u_1	2.667	1
			u_2	3.667	0.707

Six participants stated that the robot's instructions could have been more precise as an improvement. However, only two of them stated that such improved instructions should include the estimated distance between the people and the robot. When asked if they could imagine using the robot in their own homes, four *ST* and three *DT* did, four *ST* and one *DT* would do so for a more advanced prototype of the robot, and ten *ST* and 14 *DT* did not. The remaining two participants did not answer this question. The fact that the robot did not show any behavior they considered useful for them was given as the most common answer, while the fact that the robot was malfunctioning during the second phase was only mentioned by two participants. The question about their willingness to let the robot navigate in their homes on its own while they are on vacation was answered positively by five *ST* and six *DT* and negatively by five *ST* and eight *DT* participants. The remaining seven *ST* and four *DT* participants would only do so if there was a person present or when the navigation would work more robust.

As we observed that participants helped the robot regardless of *ST* or *DT* condition we argue that the role of a person had little to no impact on their helping behavior, even when the robot needed their assistance repeatedly.

RQ1: Are users repeatedly willing to help the same robot?

To understand if people are willing to assist a robot in need we measured how often the robot received help from one of the users. One of the users helped the robot in every failure situation during the *ST* condition. In the *DT* condition one of the dyads did not help the robot, even after multiple attempts of the robot to ask for help and giving instructions. In this situation both participants asked the experimenter to come back into the room, who explained that they should simply follow the instructions given by the robot, which they did for the rest of the experiment. We observed proactively given help from some participants when they noticed that the robot might ask for help. When this happened, and no instructions were actually given to them, they tried recovery behaviors they previously performed (even if these were for a different kind of failure). Our first research question *Are users repeatedly willing to help the same robot?* can be answered positively by looking at the detailed results in Table 5.3, which shows that every single participant was willing to help in cases of repeated failures of the robot.

RQ2: In a multi-user setting, who is going to help?

To answer our second research question we analyzed the data in hope of finding patterns to use for an autonomous robot to base its decisions on. We found that in both ST and DT condition 90.18% of the time the participant who gave the last command to the robot was the same to help the robot. When both participants moved to the robot to help (nine times), only once the person who commanded the robot last was not actively helping but only stood next to the robot. In one other out of these nine situations the participant who gave the last command was not the first person to assist the robot with the recovery. Our interpretation of the collected data is that a feeling of responsibility comes with the fact that a given command led to the failure of the robot. A feeling of responsibility seems to urge this specific person to act in order to aid in the recovery process.

For a mobile robot in a more open environment than our setup however, such a prediction criteria will not always be useful. The command could be given from a remote location, the robot moved to an area far away from the person who gave the command, or a task was simply scheduled autonomously. To overcome such situations we also analyzed how the distance between the robot and the participants influenced their decision-making in order to help the robot or not.

The possibility that the closest user assists more often can be helpful in situations in which (1) the last command from a user is too far back in time, (2) the needed help prevents the robot to seek this user, or (3) when the robot is acting autonomously to achieve a certain task that has not been triggered by a user. A situation when the battery charge level sinks below a given threshold is an example for such an automatically triggered task.

Figure 5.2 shows the total time between T_{ask} (the moment the robot asked for help) and the T_{end} (the time the helping ended) as t_{help}. Further t_{react}, the duration between T_{ask} and the first visible reaction T_{move} (movement of a person towards the robot) is plotted. Due to the fact that some users closely observed the robot's movement after the first failures, some reacted before the robot started to ask for help. In such situations we observed negative values for t_{react}, as the measurement was designed to be taken from the robot's question. This decision was made as we were not able to guarantee the same time span between the occurrence of the failure and the robot asking for help, thus increasing potential time variations between participants based on their attention to the robot or their given task.

Table 5.3: Number of failure situations and which user helped; u_{last} - the last user giving a command; $u_{closest}$ - the closer user helping; u_{both} - both users helping.

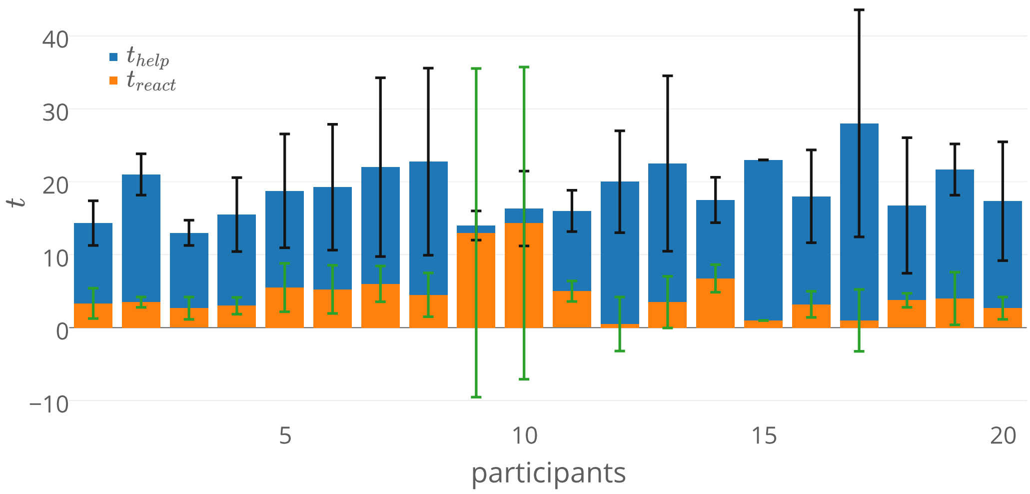
	u_{last}	$u_{closest}$	u_{both}	given help	requested help
ST	52	43	6	58	58
DT	49	46	3	54	54
$\sum$	101	89	9	112	112
percentage	90.18%	79.46%	8.04%	100%	

Algorithm 1 User selection to ask for assistance

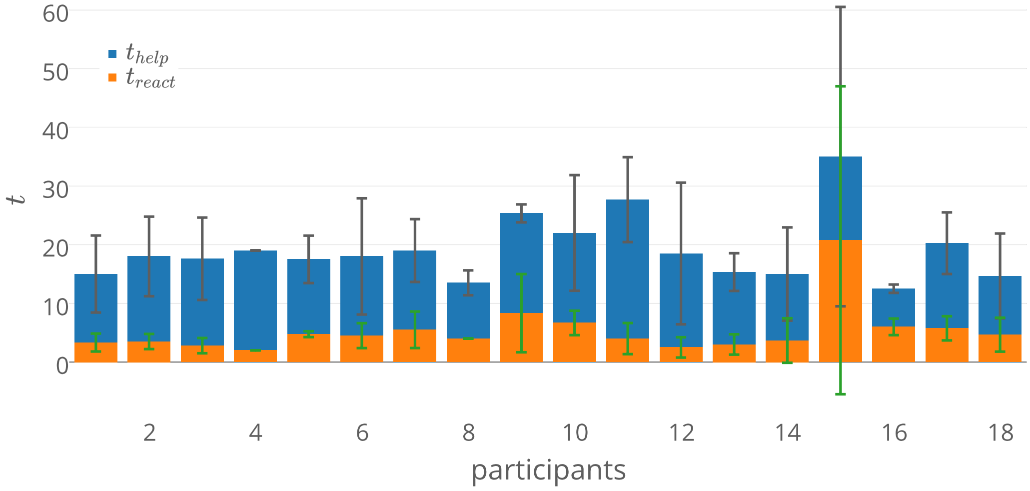
Require: failureResolved ← `False`, firstRun ← `True`
reactionThreshold ← t_{react}, helpTreshold ← t_{help}
users ← GetAllVisibleUsers(sensorData)
while ¬failureResolved $|users| > 0$ **do**
  if $t_{now} - t_{lastcmd} <$ commandTreshold firstRun **then**
    user ← GetLastCommandingUser()
    firstRun ← `False`
  else
    user ← GetClosestUser()
  end if
  if AbleToNavigateSafely() **then**
    NavigateTo(user)
  end if
  PutFocusOnUser(getHeadPose(user))
  AskForHelp(user, helpType)
  willing ← WaitForReaction(reactionThreshold)
  if willing **then**
    failureResolved ← Resolve(helpType, helpTreshold)
  end if
  if ¬failureResolved **then**
    users ← users − user
  end if
end while
return failureResolved

Implications for Autonomous HRI

Up to now Rosenthal et al. provide the only architecture that proactively incorporates the help from users in their planner [Rosenthal and Veloso [89]]. However, their algorithm is designed to deal with known limitations of the robot, i.e. the missing ability to press an elevator button. Our motivation is to develop an architecture



(a) Same Task condition (ST)



(b) Different Task condition (DT)

Figure 5.2: Average durations from the moment the robot asked for help until the participant started moving towards the robot t_{react} and until the robot acknowledged that the assistance was successful t_{help} for ST (a) and DT (b) conditions. The durations did not differ significantly between both conditions.

that allows us to predict who to ask to recover from a navigation problem. We use Algorithm 1 as a simple method to incorporate the increased willingness to help of a user that gave the last command or is otherwise closest to the robot to select the user to ask for assistance.

5.4 Conclusions

In the presented work we investigated if user support in situations where a robot is repeatedly malfunctioning can be a beneficial mitigation strategy. Our data gave us empirical evidence on users' reactions and willingness to help a robot that is supposed to assist them in fulfilling a task. This presents us the opportunity to model "planning for help" in the robot behavior for situations we cannot foresee and where the robot can only ask for help to keep the interaction alive. Moreover, the reflective self-reporting data from participants is also encouraging that even frequent malfunctioning situations do not heavily negatively impact users' perception of the robot. In other words, integrating mitigating strategies, such as the ones presented in this chapter, into the robot's behavior coordination can bring us one step further to have autonomous service robots in domestic environments in near future; especially when we know that people are willing to accept that their robot needs some help sometimes.

Chapter 6

Conclusion

The deployment of mobile service robots in private homes has emerged as a promising avenue to support older individuals living alone, offering the potential to enhance their independence and quality of life. However, for these robots to go beyond being perceived as mere tools that lose appeal over time, it is crucial to provide them with adaptive behavior, enabling extended and meaningful long-term HRI.

When the robot can seamlessly adapt its assistance to align with the user's lifestyle and expectations, it becomes more than a mere tool but a valued and integral part of the user's daily routine. This, in turn, fosters user acceptance, emotional attachment, and the long-term integration of service robots into the household.

6.1 Discussion

Returning to our research questions from Section 1.1, we can now address the following key points.

Can long-term Human-Robot Interaction be achieved on a mobile service robot using an adaptive behavior coordination system? We showed that prolonged HRI over three weeks for each of our 18 participants using the Hobbit robots was successful. While it is still an open discussion about seeing three weeks as long-term HRI, we see it as an important step to take. Our system worked well enough to push the eventual occurring wearing off due to the novelty effect, it would not be enough to keep the user interested enough for a duration of multiple months or longer.

How should a robot behave to perform tasks based on the Human in the loop principle? Maintaining a common knowledge of the state of the task and the participants, i.e. the person and the robot, needs to be achieved at any time during the interaction. Situations in which this principle was broken during our field trials left the strongest impressions on our participants. When the robot is either not able to maintain a consistent state, either internally between different software modules or externally with the user, it needs to gracefully reset to a known state to attempt to restart the ongoing task. Not addressing this issue will lead to inconsistent task executions, mismatched expectation of and the actual behavior of the robot, which in our field trials was reported as the robot making mistakes, failing to perform a task, and not informing the user of what it is currently doing.

How can the behavior of a robot be adapted to the user to improve long-term HRI? Hierarchical Finite State Machines-based behavior coordinators can only achieve a global level of adaptability if it allows to rebuild the structure of the state machines during runtime, while keeping the currently active state known. Indeed, each element of the Behavior Trees is modular, allowing it to be invoked from any other section of the tree. This flexibility enables the ability to transition to a different branch of the tree and perform alternative actions if the current evaluation determines that the current path is no longer applicable or relevant.

The decision to base our behavior coordinator on HFSM, and integrate available software modules through ROS enabled us to perform trials which went beyond the state of the art at the time. Problems with this solution arose however, when the communication between it and a module like the GUI broke down, leaving their individual states out of sync. Even though situations like that are not a unique problem in long-term HRI, repeated occurance lead to lower perceived safety, and trust into the system.

While we showed that HFSM are capable of prolonged field trials, scaling to an even larger number of behaviors would only increase the complexity of the whole behavior coordination system, which makes it at least challenging to comprehend or expand. Adaptive components within the behavior of a robot need to be implemented within the states of the HFSM as it is not possible to update the structure of the global state machine from within our implementation using the SMACH library. This means any adaptation can only occur within a specific state, not on the structural level of the behavior coordinator. To achieve such global adaptability, Behavior Trees (BT) have been gaining popularity in robotics over the last few years due to their support for task hierarchy, action sequencing, reactivity, modularity, and reusability [Styrud et al. [107]], [Colledanchise and Ögren [24]], [Shoulson et al. [101]], [Sekhavat [97]], and [Iovino et al. [56]]. Their use for complex use-cases has not yet been reported beyond 1-2 behaviors in robotics [Cooper and Lemaignan [25]].

Running trials that span multiple weeks or even longer poses practical difficulties, making such endeavors challenging or sometimes infeasible for many researchers. These trials often necessitate controlled environments, diverse participant groups, and the ability to monitor and assess the robots' performance over extended durations. As a result, researchers may find it more practical to focus on specific aspects of a robot's behavior or individual interactions to make progress within their available resources.

Nevertheless, it is vital not to lose sight of the larger picture and the overarching goals of prolonged HRI research. While researchers may concentrate on specific parts of a robot or individual interaction scenarios, it is crucial to continuously loop back the results of such studies into the overall context of the field. By doing so, researchers can contribute to the collective advancement of long-term Human-Robot Interaction.

6.2 Future work

In the context of the discussed research, we outline potential avenues for further exploration and development to advance the field of long-term Human-Robot Interaction and the deployment of service robots in private homes.

Within adaptive behavior coordination systems for long-term HRI the use of Behavior Trees due to their ability to update the active tree during runtime should be further explored. Progress in this area will allow for faster reactive behaviors and the possibility to automatically generate behaviors based on sensed state of the environment.

Conduct extensive user studies to gather in-depth insights into the needs, preferences, and expectations of older individuals living alone. Use these findings to inform the design and evaluation of service robots, ensuring their functionalities align with user requirements and promote user acceptance.

Conduct cross-cultural studies to understand how cultural backgrounds and societal norms impact user expectations, acceptance, and interactions with service robots. Explore how robots can be designed to be culturally sensitive and adaptive to different user groups.

Explore methods to enhance the robustness and reliability of service robots operating in real-world home environments over extended periods. Develop techniques for fault detection and recovery, error handling, and self-maintenance to ensure seamless and safe interactions between robots and users.

List of Tables

2.1 Comparison of related user- and field trials (including Hobbit) . . . 19

4.1 Set of hand/arm gestures recognized by Hobbit 52
4.2 Usage Frequencies of Modalities - Austrian users 57
4.3 Usage Frequencies of Modalities - Swedish users 58
4.4 Usage Frequencies of Modalities - Greek users 58
4.5 Usage Frequencies of Robot actions - Austrian users started/canceled 58
4.6 Usage Frequencies of Robot actions - Swedish users started/canceled 59
4.7 Usage Frequencies of Robot actions - Greek users started/canceled . 59
4.8 System reliability across 18 PU 64
4.9 Self-reporting of 18 users on how often each of the 17 Hobbit functions were used and how they valued the functions as working well, neither well nor badly or badly. Functions often used by most users are sorted first. 68
4.10 Self-reported NARS questionnaire scores 69
4.11 Self-reported perceived safety scores 70
4.12 Scores on the self-efficacy scale 71

5.1 Descriptive results on Perceived Intelligence (PI), Likability (LI), task contribution self (TC_S), task contribution other (TC_O), and task contribution robot (TC_R). Lower ratings could be observed for PI, LI, and TC_R in Phase 2. 89
5.2 Average number of help actions that the participants performed for the robot, broken down to every first and second user for both conditions *st* and *dt* 90
5.3 Number of failure situations and which user helped; u_{last} - the last user giving a command; $u_{closest}$ - the closer user helping; u_{both} - both users helping. 92

F.1 Which of Hobbit?s functions did you use and how often? (Interviewer: Shortly describe each function make sure, that the user knows what is meant) 132

List of Figures

1.1 Hobbit robots used by participants during field trials. Male user pointing to an object on the floor, which the robot should pick up (left, Image source: [Bajones et al. [11]]). Female user operating the robot's Graphical User Interface (right, Image source: [Frennert et al. [44]]). 3

2.1 Prototypes and mock-ups of MOVAID, URMAD, and IMMEDIATE robotic platforms 10
2.2 The robots Pearl (Nursebot project), Charles (Peoplebot), and Healthbot. 11
2.3 Care-O-Bot designs from first to fourth generation 12
2.4 A5 robot (MetraLabs) and Pepper (Aldebaran) 13
2.5 PAL Robotics TIAGo and ARI 14
2.6 Three Kompai generations 15
2.7 Paro, AIBO (the atest generation as of 2023), and Huggable robots 16

3.1 Behavior is a mapping between sensor input and motor action of actuators. 26
3.2 Hierarchical paradigm 27
3.3 Horizontal decomposition according to [Brooks [19]] 27
3.4 Behavior consists of a pair of *Sense* and *Act* stages 28
3.5 Vertical decomposition according to behavior-based paradigm . . . 28
3.6 Behavior-based paradigm consists of concurrent combinations of behaviors out of which the global behavior emerges. 29
3.7 Relation between the sensed environment, perceptual and motor scheme, and the robot's actuators. 30
3.8 Hybrid paradigm 30
3.9 Sensing in hybrid architectures 31
3.10 Hobbit's behavior coordination system 32
3.11 The diagram of the state machine of the *Learn object* task examplifies the increased complexity of Hierarchical Finite State Machines. . . . 36

4.1 From left to right: First mock-ups designed by secondary users, the first generation Hobbit prototype (PT1) and Hobbit as used during the field trials. 39
4.2 Rendering and hardware description of the second prototype of Hobbit (PT2). 40
4.3 *Robot command* interface of the Multi-Modal User Interface of Hobbit PT2. The six most often wished for functions are shown in the center of the screen. 41
4.4 Hobbit provides multi-modal input methods such as simple gestures, touch, and speech input. (a) shows the *No* gesture which can be used to answer Yes-or-No questions. In (b) a similar question is answered with *No* on the touch interface. 42
4.5 Hobbit performs the task of picking up a soft spectacle case during the trials for a user in Vienna 45
4.6 The top row show how the user is positioned relative to the robot and the fitted and tracked 3D skeleton of the user from the robot's point of view. The bottom row shows examples of the content of the robot's screen that is shown to the user during the fitness exercise. It includes instructions on how to improve based on the detected user's motions. 47
4.7 Qualitative results of the 3D skeletal model-based person detection and tracking method. 48
4.8 Lying user detection during lab tests. From left to right the acquired pair of RGB-D frames that forms the input to the plane estimation method. Two rightmost frames: plane detection and removal, and dominant outlying object detection. 50
4.9 Vision-based emergency detection of a fallen user lying on the floor. 50
4.10 Integrated human body detection and tracking and gesture recognition. 51
4.11 Overview of the trial procedure 62

5.1 Experiment's room layout with both participants (director / builder) at their workplaces and the robot in its docking station. 84
5.2 Average durations from the moment the robot asked for help until the participant started moving towards the robot t_{react} and until the robot acknowledged that the assistance was successful t_{help} for ST (a) and DT (b) conditions. 93

Acronyms

AAL Ambient Assisted Living. 16, 17, 38

ADL Activities of Daily Living. 17

ASR Automatic Speech Recognition. 16, 40, 49, 50

BCI Brain Computer Interface. 16

BT Behavior Trees. 96, 97

CCQ Clinical COPD Questionnaire. 18

CLI Command Line Interface. 85

COPD Chronic Obstructive Pulmonary Disease. 18

DT Different Task condition. 82, 83, 87–91, 93

EADL Enhanced Activities of Daily Living. 17

FoV Field of View. 44, 48, 49, 85

GRI Gesture Recognition Interface. 40

GUI Graphical User Interface. 40, 96

HFSM Hierarchical Finite State Machines. 3, 4, 36, 96

HRI Human-Robot Interaction. 1, 3, 5–7, 9, 12, 14, 16, 22, 23, 53, 54, 77, 80, 82, 92, 95–97

IADL Instrumental Activities of Daily Living. 17

LI Likability. 89

MANOVA Multivariate analysis of variance. 88

MMUI Multi-Modal User Interface. 40, 41, 43, 44, 59, 61, 66, 75, 76

MuC Mutual Care. 38, 52, 54, 62, 66, 67, 71–73

NARS Negative Attitudes Towards Robots-scale. 55, 69

PI Perceived Intelligence. 89

PU Primary Users. 38, 52, 54–56, 60, 62, 65–67, 69–73

RANSAC Random sample consensus. 49

ST Same Task condition. 82, 83, 87, 89–91, 93

SU Secondary Users. 38, 44, 55, 56, 60, 62, 65

TTS Text To Speech. 40

VoIP Voice over IP. 42

VR Virtual Reality. 18

Bibliography

[1] Jennifer Aaker, Susan Fournier, and S Adam Brasel. When good brands do bad. *Journal of Consumer research*, 31(1):1–16, 2004.

[2] Aitor Aldoma, Federico Tombari, Luigi Di Stefano, and Markus Vincze. A global hypotheses verification method for 3d object recognition. In *European conference on computer vision*, pages 511–524. Springer, 2012.

[3] Brenna D Argall, Sonia Chernova, Manuela Veloso, and Brett Browning. A survey of robot learning from demonstration. *Robotics and autonomous systems*, 57(5):469–483, 2009.

[4] Ronald C Arkin, Ronald C Arkin, et al. *Behavior-based robotics*. MIT press, 1998.

[5] Markus Bajones, Andreas Huber, Astrid Weiss, and Markus Vincze. Towards more flexible hri: How to adapt to the user? In *Workshop on Cognitive Architectures for Human-Robot Interaction at 9th ACM/IEEE International Conference on Human-Robot Interaction*. Citeseer, 2014.

[6] Markus Bajones, Daniel Wolf, Johann Prankl, and Markus Vincze. Where to look first? behaviour control for fetch-and-carry missions of service robots, 2015.

[7] Markus Bajones, Astrid Weiss, and Markus Vincze. Log data analysis of long-term household trials: Lessons learned and pitfalls. In *Proceedings of the workshop on 'The challenge (not) to go wild! Challenges and best practices to study HRI in natural interaction settings' at HRI 2016*, 2016.

[8] Markus Bajones, Astrid Weiss, and Markus Vincze. Help, anyone? a user study for modeling robotic behavior to mitigate malfunctions with the help of the user. *arXiv preprint arXiv:1606.02547*, 2016.

[9] Markus Bajones, Astrid Weiss, and Markus Vincze. Investigating the influence of culture on helping behavior towards service robots. In *Proceedings of the companion of the 2017 ACM/IEEE international conference on human-robot interaction*, pages 75–76, 2017.

[10] Markus Bajones, David Fischinger, Astrid Weiss, Daniel Wolf, Markus Vincze, Paloma de la Puente, Tobias Körtner, Markus Weninger, Konstantinos Papoutsakis, Damien Michel, Ammar Qammaz, Paschalis Panteleris, Michalis Foukarakis, Ilia Adami, Danai Ioannidi, Asterios Leonidis, Margherita Antona, Antonis Argyros, Peter Mayer, Paul Panek, Håkan Eftring, and Susanne Frennert. Hobbit: Providing fall detection and prevention for the elderly in the real world. *Journal of Robotics*, 2018, 2018. doi: 10.1155/2018/1754657. URL `https://doi.org/10.1155/2018/1754657`.

[11] Markus Bajones, David Fischinger, Astrid Weiss, Paloma De La Puente, Daniel Wolf, Markus Vincze, Tobias Körtner, Markus Weninger, Konstantinos Papoutsakis, Damien Michel, et al. Results of field trials with a mobile service robot for older adults in 16 private households. *ACM Transactions on Human-Robot Interaction (THRI)*, 9(2):1–27, 2019.

[12] Christoph Bartneck, Dana Kulić, Elizabeth Croft, and Susana Zoghbi. Measurement instruments for the anthropomorphism, animacy, likeability, perceived intelligence, and perceived safety of robots. *International journal of social robotics*, 1(1):71–81, 2009.

[13] Randall D Beer, Hillel J Chiel, and Leon S Sterling. A biological perspective on autonomous agent design. *Robotics and Autonomous Systems*, 6(1-2): 169–186, 1990.

[14] Mary Jo Bitner, Bernard H Booms, and Mary Stanfield Tetreault. The service encounter: diagnosing favorable and unfavorable incidents. *The Journal of Marketing*, pages 71–84, 1990.

[15] Indu P Bodala and Hatice Gunes. Dynamic bayesian network modelling of user affect and perceptions of a teleoperated robot coach during longitudinal mindfulness training. *arXiv preprint arXiv:2112.02017*, 2021.

[16] Jean-Pierre Briot, Thomas Meurisse, and Frédéric Peschanski. Architectural design of component-based agents: A behavior-based approach. In Rafael H. Bordini, Mehdi Dastani, Jürgen Dix, and Amal El Fallah Seghrouchni, editors, *Programming Multi-Agent Systems*, pages 71–90, Berlin, Heidelberg, 2007. Springer Berlin Heidelberg. ISBN 978-3-540-71956-4.

[17] Elizabeth Broadbent, Ngaire Kerse, Kathryn Peri, Hayley Robinson, Chandimal Jayawardena, Tony Kuo, Chandan Datta, Rebecca Stafford, Haley Butler, Pratyusha Jawalkar, et al. Benefits and problems of health-care robots in aged care settings: A comparison trial. *Australasian journal on ageing*, 35(1): 23–29, 2016.

[18] Elizabeth Broadbent, Jeff Garrett, Nicola Jepsen, Vickie Li Ogilvie, Ho Seok Ahn, Hayley Robinson, Kathryn Peri, Ngaire Kerse, Paul Rouse, Avinesh

Pillai, et al. Using robots at home to support patients with chronic obstructive pulmonary disease: pilot randomized controlled trial. *Journal of medical Internet research*, 20(2), 2018.

[19] Rodney Brooks. A robust layered control system for a mobile robot. *IEEE journal on robotics and automation*, 2(1):14–23, 1986.

[20] Hoang-Long Cao, Pablo Gómez Esteban, De Beir Albert, Ramona Simut, Greet Van de Perre, Dirk Lefeber, and Bram Vanderborght. A collaborative homeostatic-based behavior controller for social robots in human–robot interaction experiments. *International Journal of Social Robotics*, 9:675–690, 2017.

[21] Hoang-Long Cao, Greet Van de Perre, James Kennedy, Emmanuel Senft, Pablo Gómez Esteban, Albert De Beir, Ramona Simut, Tony Belpaeme, Dirk Lefeber, and Bram Vanderborght. A personalized and platform-independent behavior control system for social robots in therapy: development and applications. *IEEE Transactions on Cognitive and Developmental Systems*, 11(3): 334–346, 2018.

[22] Tom Carlson and Yiannis Demiris. Collaborative control for a robotic wheelchair: Evaluation of performance, attention, and workload. *IEEE Transactions on Systems, Man, and Cybernetics, Part B: Cybernetics*, 42(3): 876–888, 2012. ISSN 10834419. doi: 10.1109/TSMCB.2011.2181833.

[23] Michael Cashmore, Maria Fox, Derek Long, Daniele Magazzeni, Bram Ridder, Arnau Carrera, Narcis Palomeras, Natalia Hurtos, and Marc Carreras. ROSPlan: Planning in the Robot Operating System. In *Proceedings of International Conference on AI Planning and Scheduling (ICAPS)*, pages 333–341, 2015. ISBN 9781577357315.

[24] Michele Colledanchise and Petter Ögren. *Behavior trees in robotics and AI: An introduction*. CRC Press, 2018.

[25] Sara Cooper and Séverin Lemaignan. Towards using behaviour trees for long-term social robot behaviour. In *2022 17th ACM/IEEE International Conference on Human-Robot Interaction (HRI)*, pages 737–741. IEEE, 2022.

[26] Joseph Daly, Ute Leonards, and Paul Bremner. Robots in need: How patterns of emotional behavior influence willingness to help. In *Companion of the 2020 ACM/IEEE International Conference on Human-Robot Interaction*, pages 174–176, 2020.

[27] Paolo Dario, Eugenio Guglielmelli, Vincenzo Genovese, and Maurizio Toro. Robot assistants: Applications and evolution. *Robotics and autonomous systems*, 18(1-2):225–234, 1996.

[28] Kerstin Dautenhahn. Methodology & themes of human-robot interaction: A growing research field. *International Journal of Advanced Robotic Systems*, 4 (1 SPEC. ISS.):103–108, 2007. ISSN 17298806. doi: 10.5772/5702.

[29] Paloma De La Puente, Markus Bajones, Peter Einramhof, Daniel Wolf, David Fischinger, and Markus Vincze. RGB-D sensor setup for multiple tasks of home robots and experimental results. In *IEEE International Conference on Intelligent Robots and Systems*, pages 2587–2594. Institute of Electrical and Electronics Engineers Inc., 2014.

[30] Paloma de la Puente, Markus Bajones, Christian Reuther, Daniel Wolf, David Fischinger, and Markus Vincze. Robot navigation in domestic environments: Experiences using rgb-d sensors in real homes. *Journal of Intelligent & Robotic Systems*, pages 1–16, 2018.

[31] Paloma De La Puente, David Fischinger, Markus Bajones, Daniel Wolf, and Markus Vincze. Grasping objects from the floor in assistive robotics: Real world implications and lessons learned. *IEEE Access*, 7:123725–123735, 2019.

[32] Yiannis Demiris and Bassam Khadhouri. Hierarchical attentive multiple models for execution and recognition of actions. *Robotics and Autonomous Systems*, 54(5):361–369, 2006. ISSN 09218890. doi: 10.1016/j.robot.2006.02.003.

[33] Andrew Dillon. User acceptance of information technology. *Encyclopedia of human factors and ...*, 2001. URL `http://hdl.handle.net/10150/105880http://arizona.openrepository.com/arizona/handle/10150/105880`.

[34] R. Eisma, A. Dickinson, J. Goodman, A. Syme, L. Tiwari, and A. F. Newell. Early user involvement in the development of information technology-related products for older people. *Universal Access in the Information Society*, 3 (2):131–140, 2004. ISSN 1615-5289. doi: 10.1007/s10209-004-0092-z. URL `http://link.springer.com/10.1007/s10209-004-0092-z`.

[35] Abrar Fallatah, Bohkyung Chun, Sogol Balali, and Heather Knight. " would you please buy me a coffee?" how microcultures impact people's helpful actions toward robots. In *Proceedings of the 2020 ACM Designing Interactive Systems Conference*, pages 939–950, 2020.

[36] Gabor Fazekas, Andras Toth, Pierre Rumeau, Katalin Zsiga, Tamas Pilissy, and Vincent Dupurque. Cognitive-care Robot for Elderly Assistance: Preliminary Results of Tests With Users in Their Homes. In *Proceedings - AAL Forum*, pages 145–148, Eindhoven, 2012.

[37] Heike Felzmann, Kathy Murphy, Dympna Casey, and Oya Beyan. Robot-assisted care for elderly with dementia: is there a potential for genuine end-user empowerment. In *10th ACM/IEEE international conference on human–robot interaction*, 2015.

[38] Julia Fink. *Dynamics of Human-Robot Interaction in Domestic Environments*. PhD thesis, IC, Lausanne, 2014.

[39] David Fischinger, Astrid Weiss, and Markus Vincze. Learning Grasps with Topographic Features. *International Journal of Robotics Research*, 34(9): 1167–1194, 2015. ISSN 0278-3649. doi: 10.1177/0278364915577105.

[40] David Fischinger, Peter Einramhof, Konstantinos Papoutsakis, Walter Wohlkinger, Peter Mayer, Paul Panek, Stefan Hofmann, Tobias Koertner, Astrid Weiss, Antonis Argyros, and Markus Vincze. Hobbit, a care robot supporting independent living at home: First prototype and lessons learned. *Robotics and Autonomous Systems*, 75:60–78, 2016. ISSN 09218890. doi: 10.1016/j.robot.2014.09.029. URL `http://dx.doi.org/10.1016/j.robot.2014.09.029`.

[41] Terrence Fong, Charles Thorpe, and Charles Baur. Robot, asker of questions. *Robotics and Autonomous systems*, 42(3):235–243, 2003.

[42] Jodi Forlizzi and Carl Disalvo. Service Robots in the Domestic Environment: A Study of the Roomba Vacuum in the Home. *Design*, 2006:258–265, 2006. doi: 10.1145/1121241.1121286. URL `http://www.scopus.com/inward/record.url?eid=2-s2.0-33745835464{&}partnerID=40`.

[43] Michalis Foukarakis, Ilia Adami, Danae Ioannidi, Asterios Leonidis, Damien Michel, Ammar Qammaz, Konstantinos E Papoutsakis, Margherita Antona, and Antonis A Argyros. A robot-based application for physical exercise training. In *ICT4AgeingWell*, pages 45–52, 2016.

[44] Susanne Frennert, Håkan Eftring, and Britt Östlund. Case report: Implications of doing research on socially assistive robots in real homes. *International Journal of Social Robotics*, 9(3):401–415, Jun 2017. ISSN 1875-4805. doi: 10.1007/s12369-017-0396-9. URL `https://doi.org/10.1007/s12369-017-0396-9`.

[45] Razan Ghzouli, Thorsten Berger, Einar Broch Johnsen, Andrzej Wasowski, and Swaib Dragule. Behavior trees and state machines in robotics applications, 2023.

[46] Stefan Goetze, Sven Fischer, Niko Moritz, Jens-E Appell, and Frank Wallhoff. Multimodal Human-Machine Interaction for Service Robots in Home-Care Environments. In *in Proceedings of the 1st Workshop on Speech and Multimodal Interaction in Assistive Environments, Association for Computational Linguistics*, pages 1–7, 2012.

[47] Birgit Graf, Ulrich Reiser, Martin Hägele, Kathrin Mauz, and Peter Klein. Robotic home assistant care-O-bot® 3 - Product vision and innovation platform. In *Proceedings of IEEE Workshop on Advanced Robotics and its Social Impacts, ARSO*, pages 139–144, 2009. ISBN 9781424443949. doi: 10.1109/ARSO.2009.5587059.

[48] H-M Gross, H Boehme, Ch Schroeter, Steffen Müller, Alexander König, Erik Einhorn, Ch Martin, Matthias Merten, and Andreas Bley. Toomas: interactive shopping guide robots in everyday use-final implementation and experiences from long-term field trials. In *2009 IEEE/RSJ International Conference on Intelligent Robots and Systems*, pages 2005–2012. IEEE, 2009.

[49] Horst-Michael Gross, Steffen Mueller, Christof Schroeter, Michael Volkhardt, Andrea Scheidig, Klaus Debes, Katja Richter, and Nicola Doering. Robot Companion for Domestic Health Assistance: Implementation, Test and Case Study under Everyday Conditions in Private Apartments*. *IEEE/RSJ Int. Conf. on Intelligent Robots and Systems (IROS)*, pages 5992–5999, 2015. ISSN 21530866. doi: 10.1109/IROS.2015.7354230.

[50] CW Hart, JL Heskett, and WE Sasser. The profitable art of service recovery. *Harvard business review*, 68(4):148—156, 1990. ISSN 0017-8012. URL `http://europepmc.org/abstract/MED/10106796`.

[51] Marcel Heerink, Ben Kröse, Vanessa Evers, and Bob Wielinga. The influence of social presence on acceptance of a companion robot by older people, 2008.

[52] Frederik W Heger, Laura M Hiatt, Brennan Sellner, Reid Simmons, and Sanjiv Singh. Results in sliding autonomy for multi-robot spatial assembly. In *Proceedings of the 7th International Sympsium on Artificial Intelligence, Robotics, and Automation in Space*. iSAIRAS, iSAIRAS, 2005.

[53] Pamela J Hinds, Teresa L Roberts, and Hank Jones. Whose job is it anyway? a study of human-robot interaction in a collaborative task. *Human-Computer Interaction*, 19(1):151–181, 2004.

[54] Deanna Hood, Séverin Lemaignan, and Pierre Dillenbourg. The cowriter project: Teaching a robot how to write. In *Proceedings of the Tenth Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*, pages 269–269. ACM, 2015.

[55] Helge Hüttenrauch and Kerstin Severinson Eklundh. To help or not to help a service robot: Bystander intervention as a resource in human–robot collaboration. *Interaction Studies*, 7(3):455–477, 2006.

[56] Matteo Iovino, Julian Förster, Pietro Falco, Jen Jen Chung, Roland Siegwart, and Christian Smith. On the programming effort required to generate behavior trees and finite state machines for robotic applications. *arXiv preprint arXiv:2209.07392*, 2022.

[57] Chandimal Jayawardena, I-Han Kuo, Elizabeth Broadbent, and Bruce A MacDonald. Socially assistive robot healthbot: Design, implementation, and field trials. *IEEE Systems Journal*, 10(3):1056–1067, 2016.

[58] Michiel Joosse, Manja Lohse, Jose Gonzalez Perez, and Vanessa Evers. What you do is who you are: the role of task context in perceived social robot personality. In *Robotics and Automation (ICRA), 2013 IEEE International Conference on*, pages 2134–2139. IEEE, 2013.

[59] T. Kanda, H. Ishiguro, M. Imai, and T. Ono. Development and evaluation of interactive humanoid robots. *Proceedings of the IEEE*, 92(11):1839–1850, 2004. doi: 10.1109/JPROC.2004.835359.

[60] Takayuki Kanda, Masahiro Shiomi, Zenta Miyashita, Hiroshi Ishiguro, and Norihiro Hagita. An affective guide robot in a shopping mall. In *Proceedings of the 4th ACM/IEEE International Conference on Human Robot Interaction*, HRI '09, pages 173–180, New York, NY, USA, 2009. ACM. ISBN 978-1-60558-404-1. doi: 10.1145/1514095.1514127. URL `http://doi.acm.org/10.1145/1514095.1514127`.

[61] Taemie Kim and Pamela Hinds. Who should i blame? effects of autonomy and transparency on attributions in human-robot interaction. In *Robot and Human Interactive Communication, 2006. ROMAN 2006. The 15th IEEE International Symposium on*, pages 80–85. IEEE, 2006.

[62] Hema S. Koppula, Ashesh Jain, and Ashutosh Saxena. *Anticipatory Planning for Human-Robot Teams*, pages 453–470. Springer International Publishing, Cham, 2016. ISBN 978-3-319-23778-7. doi: 10.1007/978-3-319-23778-7_30. URL `https://doi.org/10.1007/978-3-319-23778-7_30`.

[63] Tobias Körtner, Alexandra Schmid, Daliah Batko-Klein, Christoph Gisinger, Andreas Huber, Lara Lammer, and Markus Vincze. How Social Robots Make Older Users Really Feel Well–A Method to Assess Users' Concepts of a Social Robotic Assistant. In *Proceedings of the International Conference on Social Robotics*, pages 138–147. Springer, 2012. doi: 10.1007/978-3-642-34103-8_14. URL `http://link.springer.com/10.1007/978-3-642-34103-8{_}14`.

[64] Tobias Körtner, Alexandra Schmid, Daliah Batko-Klein, and Christoph Gisinger. Meeting requirements of older users? robot prototype trials in a home-like environment. In *Universal Access in Human-Computer Interaction. Aging and Assistive Environments: 8th International Conference,*

UAHCI 2014, Held as Part of HCI International 2014, Heraklion, Crete, Greece, June 22-27, 2014, Proceedings, Part III 8, pages 660–671. Springer, 2014.

[65] Dimitrios Kosmopoulos, Konstantinos Papoutsakis, and Antonis A Argyros. Online segmentation and classification of modeled actions performed in the context of unmodeled ones. In *British Machine Vision Conference (BMVC)*, pages 1–12, 2014.

[66] I-Han Kuo, Elizabeth Broadbent, and Bruce MacDonald. Designing a robotic assistant for healthcare applications. In *the 7th conference of Health Informatics New Zealand, Rotorua*, 2008.

[67] Lara Lammer, Andreas Huber, Astrid Weiss, and Markus Vincze. Mutual Care: How older adults react when they should help their care robot. In *AISB2014: Proceedings of the 3rd international symposim on New Frontiers in Human-Robot interaction*, 2014. URL http://hobbit.acin.tuwien.ac.at/publications/AISB2014-HRIpaper.pdf.

[68] Karine Lan Hing Ting, Dimitri Voilmy, Quitterie De Roll, Ana Iglesias, and Rebeca Marfil. Fieldwork and field trials in hospitals: Co-designing a robotic solution to support data collection in geriatric assessment. *Applied Sciences*, 11(7):3046, 2021.

[69] Min Kyung Lee, Sara Kielser, Jodi Forlizzi, Siddhartha Srinivasa, and Paul Rybski. Gracefully mitigating breakdowns in robotic services. In *Proceedings of the 5th ACM/IEEE International Conference on Human-robot Interaction*, HRI '10, pages 203–210, Piscataway, NJ, USA, 2010. IEEE Press. ISBN 978-1-4244-4893-7. URL http://dl.acm.org/citation.cfm?id=1734454.1734544.

[70] Iolanda Leite, Carlos Martinho, and Ana Paiva. Social Robots for Long-Term Interaction: A Survey. *International Journal of Social Robotics*, 5(2):291–308, 2013. ISSN 18754791. doi: 10.1007/s12369-013-0178-y.

[71] Dan Li, Dingwei Gao, Suyun Fan, GangHua Lu, Wen Jiang, Xueyu Yuan, Yanyan Jia, Ming Sun, Jianjun Liu, Zairong Gao, et al. Effectiveness of mobile robots collecting vital signs and radiation dose rate for patients receiving iodine-131 radiotherapy: A randomized clinical trial. *Frontiers in Public Health*, 10:1042604, 2023.

[72] Masoumeh Mansouri and Federico Pecora. More knowledge on the table: Planning with space, time and resources for robots. In *Proceedings - IEEE International Conference on Robotics and Automation*, pages 647–654, 2014. ISBN 978-1-4799-3685-4. doi: 10.1109/ICRA.2014.6906923.

[73] J. C. Marquié, L. Jourdan-Boddaert, and N. Huet. Do older adults underestimate their actual computer knowledge? *Behaviour and Information Technology*, 21(4):273–280, 2002. ISSN 0144929X. doi: 10.1080/0144929021000020998.

[74] Richard A McFarland. *Physiological psychology: The biology of human behavior*. Mayfield Publishing Company, 1981.

[75] D Michel and A Argyros. Apparatuses, methods and systems for recovering a 3-dimensional skeletal model of the human body, 2016. URL `http://www.freepatentsonline.com/y2016/0086350.html`.

[76] D Michel and K Papoutsakis. Gesture Recognition Apparatuses, Methods and Systems for Human-Machine Interaction, 2016. URL `http://www.freepatentsonline.com/y2016/0078289.html`.

[77] Damien Michel, Kostas Papoutsakis, and Antonis Argyros. Gesture recognition supporting the interaction of humans with socially assistive robots. *Advances in Visual Computing*, 2014. URL `http://link.springer.com/chapter/10.1007/978-3-319-14249-4{_}76`.

[78] Stanislava Naneva, Marina Sarda Gou, Thomas L Webb, and Tony J Prescott. A systematic review of attitudes, anxiety, acceptance, and trust towards social robots. *International Journal of Social Robotics*, 12(6):1179–1201, 2020.

[79] Monica Nicolette Nicolescu. *A framework for learning from demonstration, generalization and practice in human-robot domains*. PhD thesis, University of Southern California, 2003.

[80] Nils J Nilsson. Shakey the robot. Technical report, SRI INTERNATIONAL MENLO PARK CA, 1984.

[81] Tatsuya Nomura, Tomohiro Suzuki, Takayuki Kanda, and Kensuke Kato. Measurement of negative attitudes toward robots. *Interaction Studies*, 7(3):437–454, 2006. ISSN 15720373. doi: 10.1075/is.7.3.14nom. URL `http://openurl.ingenta.com/content/xref?genre=article{&}issn=1572-0373{&}volume=7{&}issue=3{&}spage=437`.

[82] Filippo Palumbo, Davide La Rosa, Erina Ferro, Davide Bacciu, Claudio Gallicchio, Alessio Micheli, Stefano Chessa, Federico Vozzi, and Oberdan Parodi. Reliability and human factors in ambient assisted living environments: The doremi case study. *Journal of Reliable Intelligent Environments*, 3:139–157, 2017.

[83] Martha E Pollack, Laura Brown, Dirk Colbry, Cheryl Orosz, Bart Peintner, Sailesh Ramakrishnan, Sandra Engberg, Judith T Matthews, Jacqueline Dunbar-Jacob, Colleen E McCarthy, et al. Pearl: A mobile robotic assistant

for the elderly. In *AAAI workshop on automation as eldercare*, volume 2002, pages 85–91, 2002.

[84] Johann Prankl, Aitor Aldoma, Alexander Svejda, and Markus Vincze. Rgb-d object modelling for object recognition and tracking. In *Intelligent Robots and Systems (IROS), 2015 IEEE/RSJ International Conference on*, pages 96–103. IEEE, 2015.

[85] Jürgen Pripfl, Tobias Körtner, Daliah Batko-Klein, Denise Hebesberger, Markus Weninger, Christoph Gisinger, Susanne Frennert, Hakan Eftring, Margarita Antona, Ilia Adami, Astrid Weiss, Markus Bajones, and Markus Vincze. Results of a real world trial with a mobile social service robot for older adults. In *Proceedings of the Eleventh Annual ACM/IEEE International Conference on Human-Robot Interaction Extended Abstracts*, 2016.

[86] Mehdi Ravangard. Fuzzy behavior based mobile robot navigation. In *2015 4th Iranian Joint Congress on Fuzzy and Intelligent Systems (CFIS)*, pages 1–7. IEEE, 2015.

[87] Tobias Rehrl, Raphaël Troncy, Andreas Bley, Susanne Ihsen, Katharina Scheibl, Sebastian Glende, Stefan Goetze, Jens Kessler, Christoph Hintermueller, and Frank Wallhoff. The Ambient Adaptable Living Assistant is Meeting its Users. *Proceedings of the AAL Forum 2012*, pages 629–636, 2012. doi: 10.1.1.364.3969.

[88] Yvonne Rogers. Interaction design gone wild. *interactions*, 18(4):58, jul 2011. ISSN 10725520. doi: 10.1145/1978822.1978834. URL `http://doi.acm.org/10.1145/1978822.1978834http://portal.acm.org/citation.cfm?doid=1978822.1978834`.

[89] Stephanie Rosenthal and Manuela M Veloso. Mobile robot planning to seek help with spatially-situated tasks. *AAAI*, 4(5.3):1, 2012.

[90] Stephanie Rosenthal, Joydeep Biswas, and Manuela Veloso. An effective personal mobile robot agent through symbiotic human-robot interaction. In *Proceedings of the 9th International Conference on Autonomous Agents and Multiagent Systems: volume 1-Volume 1*, pages 915–922. International Foundation for Autonomous Agents and Multiagent Systems, 2010.

[91] Astrid Marieke Rosenthal-von der Pütten, Astrid Weiss, and Selma Šabanović. The challenge (not) to go wild!: Challenges and best practices to study hri in natural interaction settings. In *The Eleventh ACM/IEEE International Conference on Human Robot Interaction*, HRI '16, pages 583–584, Piscataway, NJ, USA, 2016. IEEE Press. ISBN 978-1-4673-8370-7. URL `http://dl.acm.org/citation.cfm?id=2906831.2906991`.

[92] Selma Sabanovic, Marek P. Michalowski, and Reid Simmons. Robots in the wild: Observing human-robot social interaction outside the lab. In *International Workshop on Advanced Motion Control, AMC*, volume 2006, pages 576–581, 2006. ISBN 0-7803-9511-1. doi: 10.1109/AMC.2006.1631758.

[93] Alessandra Maria Sabelli and Takayuki Kanda. Robovie as a mascot: a qualitative study for long-term presence of robots in a shopping mall. *International Journal of Social Robotics*, 8:211–221, 2016.

[94] Richard A Schmidt. A schema theory of discrete motor skill learning. *Psychological review*, 82(4):225, 1975.

[95] Ch Schroeter, S. Mueller, M. Volkhardt, E. Einhorn, C. Huijnen, H. Van Den Heuvel, A. Van Berlo, A. Bley, and H. M. Gross. Realization and user evaluation of a companion robot for people with mild cognitive impairments. *Proceedings - IEEE International Conference on Robotics and Automation*, pages 1153–1159, 2013. ISSN 10504729. doi: 10.1109/ICRA.2013.6630717.

[96] Ralf Schwarzer and Matthias Jerusalem. Generalized self-efficacy scale. *J. Weinman, S. Wright, & M. Johnston, Measures in health psychology: A user's portfolio. Causal and control beliefs*, 35:37, 1995.

[97] Yoones A Sekhavat. Behavior trees for computer games. *International Journal on Artificial Intelligence Tools*, 26(02):1730001, 2017.

[98] T. Shibata and K. Tanie. Physical and affective interaction between human and mental commit robot. In *Proceedings 2001 ICRA. IEEE International Conference on Robotics and Automation (Cat. No.01CH37164)*, volume 3, pages 2572–2577 vol.3, 2001. doi: 10.1109/ROBOT.2001.933010.

[99] Masahiro Shiomi, Takayuki Kanda, Hiroshi Ishiguro, and Norihiro Hagita. Interactive humanoid robots for a science museum. In *Proceedings of the 1st ACM SIGCHI/SIGART Conference on Human-robot Interaction*, HRI '06, pages 305–312, New York, NY, USA, 2006. ACM. ISBN 1-59593-294-1. doi: 10.1145/1121241.1121293. URL `http://doi.acm.org/10.1145/1121241.1121293`.

[100] Masahiro Shiomi, Daisuke Sakamoto, Takayuki Kanda, Carlos Toshinori Ishi, Hiroshi Ishiguro, and Norihiro Hagita. A semi-autonomous communication robot: a field trial at a train station. In *Proceedings of the 3rd ACM/IEEE international conference on Human robot interaction*, pages 303–310. ACM, 2008.

[101] Alexander Shoulson, Francisco M Garcia, Matthew Jones, Robert Mead, and Norman I Badler. Parameterizing behavior trees. In *Motion in Games: 4th International Conference, MIG 2011, Edinburgh, UK, November 13-15, 2011. Proceedings 4*, pages 144–155. Springer, 2011.

[102] Cory-Ann Smarr, Akanksha Prakash, Jenay M Beer, Tracy L Mitzner, Charles C Kemp, and Wendy 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*, volume 56, pages 153–157. Sage Publications Sage CA: Los Angeles, CA, 2012.

[103] Richard A Spreng, Gilbert D Harrell, and Robert D Mackoy. Service recovery: impact on satisfaction and intentions. *Journal of Services Marketing*, 9(1): 15–23, 1995.

[104] Rebecca Stafford. *Stafford 2013 PhD thesis - The contributions of peoples attitudes and perceptions to the acceptance of eldercare robots*. PhD thesis, University of Auckland, 10 2013.

[105] Walter Dan Stiehl, Cynthia Breazeal, Kuk-Hyun Han, Jeff Lieberman, Levi Lalla, Allan Maymin, Jonathan Salinas, Daniel Fuentes, Robert Toscano, Cheng Hau Tong, et al. The huggable: a therapeutic robotic companion for relational, affective touch. In *ACM SIGGRAPH 2006 emerging technologies*, page 15. ACM, 2006.

[106] Tilo Strutz. *Data fitting and uncertainty: A practical introduction to weighted least squares and beyond*. Springer, 2011.

[107] Jonathan Styrud, Matteo Iovino, Mikael Norrlöf, Mårten Björkman, and Christian Smith. Combining planning and learning of behavior trees for robotic assembly. In *2022 International Conference on Robotics and Automation (ICRA)*, pages 11511–11517. IEEE, 2022.

[108] JaYoung Sung, Rebecca E. Grinter, and Henrik I. Christensen. "pimp my roomba": Designing for personalization. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '09, page 193–196, New York, NY, USA, 2009. Association for Computing Machinery. ISBN 9781605582467. doi: 10.1145/1518701.1518732. URL `https://doi.org/10.1145/1518701.1518732`.

[109] Vincent Wing Sun Tung and Norman Au. Exploring customer experiences with robotics in hospitality. *International Journal of Contemporary Hospitality Management*, 2018.

[110] Michel Vacher, François Portet, Anthony Fleury, and Norbert Noury. Development of Audio Sensing Technology for Ambient Assisted Living. *International Journal of E-Health and Medical Communications*, 2(1):35–54, 2012. ISSN 1947-315X. doi: 10.4018/jehmc.2011010103. URL `https://hal.archives-ouvertes.fr/hal-00757407`.

[111] Andrea Vanzo, Francesco Riccio, Mahmoud Sharf, Valeria Mirabella, Tiziana Catarci, and Daniele Nardi. Who is willing to help robots? a user study on collaboration attitude. *International Journal of Social Robotics*, 12:589–598, 2020.

[112] Markus Vincze and Markus Bajones. What a year of trials with a mobile robot in user homes reveals about the actual user needs. In *Workshop on the Barriers of Social Robotics Take-Up by Society, held at the 26th IEEE International Symposium on Robot and Human Interactive Communication. IEEE, Leicester*, 2017.

[113] Markus Vincze, Markus Bajones, Markus Suchi, Daniel Wolf, Astrid Weiss, David Fischinger, and Paloma da la Puente. Learning and detecting objects with a mobile robot to assist older adults in their homes. In *European Conference on Computer Vision*, pages 316–330. Springer, 2016.

[114] Markus Vincze, David Fischinger, Markus Bajones, Daniel Wolf, Markus Suchi, Lara Lammer, Astrid Weiss, Juergen Pripfl, Tobias Koertner, and Christoph Gisinger. What older adults would like a robot to do in their homes-first results from a user study in the homes of users. In *ISR 2016: 47st International Symposium on Robotics; Proceedings of*, pages 1–7. VDE, 2016.

[115] Paul Viola and Michael J MJ Jones. Robust Real-Time Face Detection. *International Journal of Computer Vision*, 57(2):137–154, 2004. ISSN 09205691. URL http://link.springer.com/article/10.1023/B:VISI.0000013087.49260.fb.

[116] K. Wada and T. Shibata. Living with seal robots - its sociopsychological and physiological influences on the elderly at a care house. *IEEE Transactions on Robotics*, 23(5):972–980, Oct 2007. ISSN 1552-3098. doi: 10.1109/TRO.2007.906261.

[117] K. Wada, T. Shibata, T. Saito, and K. Tanie. Effects of robot-assisted activity for elderly people and nurses at a day service center. *Proceedings of the IEEE*, 92(11):1780–1788, Nov 2004. ISSN 0018-9219. doi: 10.1109/JPROC.2004.835378.

[118] K. Wada, T. Shibata, T. Saito, K. Sakamoto, and K. Tanie. Psychological and social effects of one year robot assisted activity on elderly people at a health service facility for the aged. In *Proceedings of the 2005 IEEE International Conference on Robotics and Automation*, pages 2785–2790, April 2005. doi: 10.1109/ROBOT.2005.1570535.

[119] Astrid Weiss, Judith Igelsböck, Manfred Tscheligi, Andrea Bauer, Kolja Kühnlenz, Dirk Wollherr, and Martin Buss. Robots asking for directions: the willingness of passers-by to support robots. In *Proceedings of the 5th*

ACM/IEEE international conference on Human-robot interaction, pages 23–30. IEEE Press, 2010.

[120] Astrid Weiss, Betsy van Dijk, and Vanessa Evers. Knowing me knowing you: Exploring effects of culture and context on perception of robot personality. In *Proceedings of the 4th international conference on Intercultural Collaboration*, pages 133–136. ACM, 2012.

[121] Astrid Weiss, Nicole Mirnig, Ulrike Bruckenberger, Ewald Strasser, Manfred Tscheligi, Dirk Wollherr, Bartlomiej Stanczyk, et al. The interactive urban robot: User-centered development and final field trial of a direction requesting robot. *Paladyn, Journal of Behavioral Robotics*, 6(1), 2015.

[122] Daniel Wolf, Markus Bajones, Johann Prankl, and Markus Vincze. Find my mug: Efficient object search with a mobile robot using semantic segmentation. *arXiv preprint arXiv:1404.5765*, 2014.

[123] Murphy Wonsick and Taskin Padir. A systematic review of virtual reality interfaces for controlling and interacting with robots. *Applied Sciences*, 10 (24):9051, 2020.

[124] Gabriele Wulf. *Motor Schema*, pages 2350–2352. Springer US, Boston, MA, 2012. ISBN 978-1-4419-1428-6. doi: 10.1007/978-1-4419-1428-6_870. URL `https://doi.org/10.1007/978-1-4419-1428-6_870`.

[125] Lucy Yardley, Nina Beyer, Klaus Hauer, Gertrudis Kempen, Chantal Piot-Ziegler, and Chris Todd. Development and initial validation of the Falls Efficacy Scale-International (FES-I). *Age and Ageing*, 34(6):614–619, 2005. ISSN 00020729. doi: 10.1093/ageing/afi196.

[126] Katalin Zsiga, András Tóth, Tamás Pilissy, Orsolya Péter, Zoltán Dénes, and Gábor Fazekas. Evaluation of a companion robot based on field tests with single older adults in their homes. *Assistive Technology*, 30(5):259–266, 2018. doi: 10.1080/10400435.2017.1322158. URL `https://doi.org/10.1080/10400435.2017.1322158`. PMID: 28628395.

Appendices

Appendix A

Interview guidelines

1. Explain the purpose/goal of the Interview (we would like to get an impression of the time you spent together with HOBBIT and you experiences)
2. Make participants aware that the questions refer to the robot and not the (potential influence/presence) of a facilitator or technician.
3. Number of questions
4. Approximate duration of the interview
5. Informing the elderly that the interview is sound recorded and asked for consent
6. Informing the elderly that the data will be anonymized.
7. Does the interview partner have any questions in regard of the interview, procedure etc.

Appendix B

Timeline

- Screening interview
  - Screening Questionnaire
  - Sight
  - Hearing
  - Mobility
- Pre-phase
  - Falls Efficacy Scale (shortened version)
  - Ethics and Attachment (questions 1 to 14)
  - Self Efficacy Scale
  - NARS
- Midtrial assessment
  - Falls Efficacy Scale (shortened version)
  - Ethics and Attachment (questions 1 to 14)
  - Self Efficacy Scale
  - NARS
  - Midtrial interview
- End-of-trial assessment
  - Falls Efficacy Scale (shortened version)
  - Ethics and Attachment (questions 1 to 14)
  - Self Efficacy Scale

  - NARS
  - End-of-trial interview - Primary User
  - End-of-trial interview - Secondary User

- Post-phase
  - Falls Efficacy Scale (shortened version)
  - Ethics and Attachment (questions 1 to 15)
  - Self Efficacy Scale
  - NARS
  - Post-phase interview

Appendix C

Screening Questionnaire

1. Age . . . . [ ]
2. Sex . . . . [female] [male]
3. Handedness . . . . [left-handed] [right-handed]
4. What is your highest level of education? P=Primary school, L=Lower secondary education, V=Vocational school, A=Apprenticeship, H=Higher secondary education, U=University degree . . . . [P] [L] [V] [A] [H] [U]
5. What was your former occupation? W=Worker, S=Self employed, E=Employee, L=Leading position, H=Home keeper, U=Unemployed . . . . [W] [S] [E] [L] [H] [U]
6. What is your living situation? . . . . [Flat] [House] [Assisted living]
7. How often do you use a computer? N=never/less than once a week, O=once a week, T=two or three times a week, M=more often than three times a week, D=every day . . . . [N] [O] [T] [M] [D]
8. Do you have a cell-phone? . . . . [Yes] [No]

▸ 9. If Yes: What kind of phone?

____

____

10. Do you sometimes feel dizzy? . . . . [Yes] [No]

▸ 11. Further information

12. Have you ever fallen at home and needed help during the last two years? . . . . Yes No

▸ 13. Further information

14. Are you generally worried or afraid that you might fall? Yes No

Possible answers: M=Mostly, S=Sometimes, N=Never, NR=Not relevant

Sight

1. I use glasses.	M	S	N	NR
2. I am capable to read a newspaper without difficulties. .	M	S	N	NR
3. I am capable to recognize a face from 4 meters distance.	M	S	N	NR
4. I find it difficult to read and use buttons of a regular phone or remote control.	M	S	N	NR
5. I find it difficult to read regular gas, electricity and telephone bills or bank statements.	M	S	N	NR
6. I find it difficult to distinguish between a 1 and a 2 coin, or between a 10cent and a 20cent coin.	M	S	N	NR

Hearing

1. I am capable to hear a conversation without difficulties when several people are talking (without using a hearing aid).	M	S	N	NR
2. I am capable to hear a conversation with one person without difficulties (without using a hearing aid).	M	S	N	NR
3. I am capable to hear what other people say to me without difficulties (using a hearing aid).	M	S	N	NR
4. I find it easy to listen to the voice of a young kid (high/shrill voice).	M	S	N	NR

5. I find it difficult to listen to someone speaking on the phone.	M	S	N	NR
6. People around me complain about the volume of my television set.	M	S	N	NR

Mobility

1. I am capable to walk 500 meters carrying a bag of 5kg without support	M	S	N	NR
2. I shop on my own and carry the grocery back home by myself.	M	S	N	NR
3. I am capable to stoop, bend or kneel without difficulties.	M	S	N	NR
4. I am capable to reach my hands over my head without difficulties.	M	S	N	NR
5. I can easily walk up two floors.	M	S	N	NR
6. I am afraid of falling in my house.	M	S	N	NR

Appendix D

Pre-phase interview

Falls efficacy scale(shortened version)
Below are some questions about how concerned you are about the possibility of falling. Please reply thinking about how you usually do the activity. If you currently don?t do the activity (for example, if someone does your shopping for you), please answer to show whether you think you would be concerned about falling IF you did the activity. For each of the following activities, please check the box which is closest to your own opinion to show how concerned you are that you might fall if you did this activity.

Possible answers:
N=Not at all concerned, S=Somewhat concerned, F=Fairly concerned, V=Very concerned

1. Cleaning the house.	N	S	F	V
2. Getting dressed or undressed.	N	S	F	V
3. Preparing simple means.	N	S	F	V
4. Taking a bath or shower.	N	S	F	V
5. Getting in or out of chair.	N	S	F	V
6. Reaching for something above your head or on the ground.	N	S	F	V
7. Going to answer the telephone before it stops ringing.	N	S	F	V

Ethics and attachment
Possible answers:
N=Not at all concerned, S=Somewhat concerned, F=Fairly concerned, V=Very

concerned

1. For me, the robot would be...	N	S	F	V
2. I suppose, I would support the robot in its tasks.	N	S	F	V
3. I suppose, I would be supported by the robot.	N	S	F	V
4. I suppose, the robot and I would support each other.	N	S	F	V
5. I suppose, it would be important for me to help the robot when needed, because it would help me too.	N	S	F	V
6. I suppose, I would take the robots needs into account as if it would be a good friend.	N	S	F	V
7. I suppose, the robot would know right from wrong.	N	S	F	V
8. I suppose, the robot would has a will of its own.	N	S	F	V
9. I suppose, I could empathize with the robot.	N	S	F	V
10. I suppose, the robot would understand what is going on in my mind.	N	S	F	V
11. I suppose, I can rely on the robot doing what it is meant for.	N	S	F	V
12. I suppose, the robot would become important to me.	N	S	F	V
13. I suppose, the robot would be like a friend.	N	S	F	V
14. I suppose, the robot would be like a pet.	N	S	F	V
15. I miss the robot.	N	S	F	V

Self efficacy scale

Possible answers:
N=Not at all true, H=Hardly true, M=Moderately true , E=Exactly true

1. I can always manage to solve difficult problems if I try hard enough.	N	H	M	E
2. If someone opposes me, I can find the means and ways to get what I want.	N	H	M	E
3. It is easy for me to stick to my aims and accomplish my goals.	N	H	M	E
4. I am confident that I could deal efficiently with unexpected events.	N	H	M	E

5. Thanks to my resourcefulness, I know how to handle unforeseen situations.	N	H	M	E
6. I can solve most problems if I invest the necessary effort.	N	H	M	E
7. I can remain calm when facing difficulties because I can rely on my coping abilities.	N	H	M	E
8. When I am confronted with a problem, I can usually find several solutions.	N	H	M	E
9. If I am in trouble, I can usually think of a solution.	N	H	M	E
10. I can usually handle whatever comes my way.	N	H	M	E

NARS

Please indicate in which way you agree or disagree to the following statements: Possible answers: DAA=I don't agree at all, D=I don't agree, U=undecided, A=I agree, TA=I totally agree

1. I would not feel comfortable if robots really had emotions.	DA	D	U	A	TA
2. Something bad might happen if robots developed into living beings.	DA	D	U	A	TA
3. I would feel relaxed talking with robots.	DA	D	U	A	TA
4. I would not feel comfortable if I was given a job where I had to use robots.	DA	D	U	A	TA
5. If robots had emotions I would be able to make friends with them.	DA	D	U	A	TA
6. I feel comforted being with robots that have emotions.	DA	D	U	A	TA
7. The word ?robot? means nothing to me.	DA	D	U	A	TA
8. I would feel nervous operating a robot in front of other people.	DA	D	U	A	TA
9. I would not like the idea that robots or artificial intelligences were making judgements about things.	DA	D	U	A	TA
10. I would feel nervous just standing in front of a robot.	DA	D	U	A	TA
11. I feel that if I depend on robots too much, something bad might happen.	DA	D	U	A	TA
12. I would feel paranoid talking with a robot.	DA	D	U	A	TA

13. I am worried that robots would be of bad influence on children.	DA	D	U	A	TA
14. I feel that in the future society will be dominated by robots.	DA	D	U	A	TA

Appendix E

Midtrial interview

1. How did you perceive the robots behavior?

2. If you would describe your HOBBIT, how would you characterize it?

Appendix F

End-of-trial interview - Primary User

1. **Utility** Does the robot offer the right functionalities, therefore older adults feel supported to maintain independent living at home?

2. **Utility** What do you need, so that you can feel independent at home?

3. **Utility** Did HOBBIT influence your feeling of independence and how?

4. **Utility** How did the presence of HOBBIT influence your daily life?

5. **Utility** What did you use HOBBIT for?

6. **Utility** In what respect did HOBBIT help you the most?

7. **Utility** Which of Hobbit?s functions did you use and how often? (Interviewer: Shortly describe each function to make sure, that the user knows what is meant)

8. **Utility** Which functions were the most important for you?

9. **Utility** Have you been satisfied with HOBBIT?S performance?

10. **Utility** Has HOBBIT always found his way and place?

11. **Utility** In your opinion, did the robot lack any functionalities? If yes, which?

Table F.1: Which of Hobbit?s functions did you use and how often? (Interviewer: Shortly describe each function make sure, that the user knows what is meant)

Function	I used this function never / sometimes / often	This function worked good / bad
Call Hobbit		
Emergency		
Teach object		
Pickup object		
Bring object		
Calendar		
Follow me		
Go to		
Go recharging		
Break		
Telephone		
Information		
Surprise me		
Entertainment (games, audio, video)		
Reward Hobbit		
Make Hobbit stop		
Come closer		

12. **Flexibility** How did you mostly communicate with the robot (screen, voice, gesture) ? and why?

___

___

13. **Flexibility** Did any problems occur when you were communicating with HOBBIT? If yes, which? What did you do to solve the problem?

___

___

14. **Ease of Learning** How did you experience communicating with HOBBIT during the first days of the trial?

15. **Ease of Learning** Was it possible for you to tell the robot what you wanted? Why?

16. **Ease of Learning** Was it possible for you to understand answers and reactions of the robot? Why?

17. **Self-efficacy** Could HOBBIT help you in situations, which would have been difficult to master on your own? If yes, which?

18. **Emotional Attachment** In your opinion, did HOBBIT show emotions? If yes, how did the robot do that and which emotions were shown?

19. **Emotional Attachment** Can you remember a situation with HOBBIT which was very joyful? (Please describe the situation)

20. **Emotional Attachment** Can you remember a situation with HOBBIT which was unpleasant? (Please describe the situation)

____

____

21. **Emotional Attachment** How did you feel when HOBBIT managed to fulfill a certain task?

____

____

22. **Emotional Attachment** How did you feel when HOBBIT did not manage to fulfill a certain task?

____

____

23. **Emotional Attachment** Did you reward your robot? If so, in which situations?

____

____

24. **Emotional Attachment** How did you feel when HOBBIT asked you for help in situations he could not manage on his own?

____

____

25. **Emotional Attachment** If you would describe your HOBBIT, how would you characterize it?

____

____

26. **Perceived Safety** Does the perceived safety change over time due to having HOBBIT in the household?

27. **Perceived Safety** What do you need to feel safe at home?

28. **Perceived Safety** Did HOBBIT influence your perception of safety? If yes, how?

29. **Perceived Safety** Do you think HOBBIT is a safe device to live with?

30. **Costumer perceived value** Do the users perceive the robot as worth buying with the PT2 configuration?

31. **Costumer perceived value** Who would you recommend to buy a robot like HOBBIT?

32. **Costumer perceived value** Would you like to always have a HOBBIT in your home? Why?

33. **Costumer perceived value** If HOBBIT would be buy-able / rent-able, would you yourself buy / rent one? Why? Why not?

---

34. **Costumer perceived value** If you had a robot in your home, do you think your social life would change? Why?

---

35. **Mutual Care Concept - Companion - Device** How did you perceive the robots behavior?

---

36. **Mutual Care Concept - Companion - Device** Did you recognize any differences in the behavior of Hobbit during the last 10 days compared to the first period, i.e. before our last interview?

---

37. **Mutual Care Concept - Companion - Device** If yes, which differences did you recognize?

▸ 38. If yes, which differences did you recognize?

- 39. If *reciprocity, i.e. return of favour* is not mentioned by the user ask: *Did you recognize that Hobbit offered you to return favours?*

- 40. If yes, do you think this is a nice feature of Hobbit which should be kept if Hobbit will be developed further?

- 41. If *amicable dialogue* is not mentioned ask: *Did you experience Hobbit being friendlier during the last 10 days compared with the first period?*

- 42. If yes, do you think this is a nice characteristic of Hobbit which should be kept if Hobbit will be developed further?

- 43. If *increased interaction or presence* is not mentioned, ask: *Did you experience Hobbit being more present, i.e. coming more often to you to ask something or staying longer in the room with you?*

- 44. If yes, do you think this is a nice characteristic of Hobbit which should be kept if Hobbit will be developed further?

45. **Mutual Care Concept - Companion - Device** Do you think that Hobbit behaved more like a friend than a machine or more like a machine than a friend during the last 10 days?

46. **Mutual Care Concept - Companion - Device** What kind of behavior would you prefer? A machine-like Hobbit or a Hobbit who behaves like a friend? Why?

47. **Closure** When you think back about your time with HOBBIT, is there anything else that you would like to tell us?

Appendix G

End-of-trial interview - Secondary User

1. **Usability** Did you help your relative to handle the robot during the last three weeks? If so ? in what aspects did your relative require your help? How often was your help required?

2. **Usability** How would you say did your relative manage to interact with the robot?

3. **Usability** How did HOBBIT affect the daily life of your relative?

4. **Self-efficacy** Is the self-efficacy maintained on the same level due to having HOBBIT in the household?

5. **Self-efficacy** How often did your relative contact you because of questions about HOBBIT or because there were some problems?

6. **Usability** Do you think, that the daily activity-pattern of your relative changed during the trial time?

7. **Perceived Safety** Does the perceived safety change over time due to having HOBBIT in the household?

8. **Perceived Safety** What do you need to have the feeling that your relative is safe in his/her home?

9. **Perceived Safety** Do you think that the presence of HOBBIT did (in this regard) change your perception of safety? If yes/no, why?

10. **Perceived Safety** How did you feel about your relative living with a robot for three weeks?

11. **Perceived Safety** Do you think HOBBIT is a safe device for serving in the home of an older person?

12. **Costumer perceived value** Do the users perceive the robot as worth buying with the PT2 configuration?

13. **Costumer perceived value** If a robot like HOBBIT costs 15.000 ?, would you buy a PT2 for your relative? Why (not)?

14. **Costumer perceived value** If it would cost 420 ? per month, would you rent a PT2 for your relative? Why (not)?

15. **Costumer perceived value** If your relative had a robot like HOBBIT, do you think it would influence your interaction with him/her? (Why? In which way?)

16. **Closure** When you think back about your time that your relative spent with HOBBIT, is there anything else that you would like to tell us?

Appendix H

Post-phase interview

A picture of HOBBIT (ideally one in the home of that particular user) is shown to the elderly person. This picture should induce memories of the time spent with the robot.

1. When you look at the picture and remember the time you spent together with HOBBIT, what comes to your mind or what do you feel?

If the user has difficulties to retrieve memories further questions could be asked:

2. How did you perceive the presence of the robot in your home?

3. What situations with the robot do you remember most?

4. What did you like most during the trials?

5. What did you not like?

6. **Utility** Does the robot offer the right functionalities, so that older adults feel supported to maintain independent living at home?

7. **Utility** You have been living with HOBBIT for 3 weeks now. Are there situations, when you wish HOBBIT would be here again? If so, which?

8. **Utility** Did your daily life change since HOBBIT has moved out again?

9. **Emotional Attachment** Do the users develop an emotional bonding towards the HOBBIT robot over time?

10. **Emotional Attachment** How do you feel, now the HOBBIT is gone?

11. **Emotional Attachment** Do you miss the HOBBIT? Why/What do you miss?

12. **Perceived Safety** Does the perceived safety change over time due to having HOBBIT in the household?

13. **Perceived Safety** We were talking about safety during the last interview ? how do you feel in regard of safety, now that you live on your own again?

14. **Mutual Care** Did you note a change in the robot's behavior between the first and second half of its stay in your home? If yes, how did it change?

15. **Closure** When you think back about your time with HOBBIT, is there anything else that you would like to tell us?