Augmented Reality at the Workplace:

A Context-Aware Assistive System using In-Situ Projection

Markus Funk

Universität Stuttgart

Augmented Reality at the Workplace

A Context-Aware Assistive System using In-Situ Projection

Von der Fakultät für Informatik, Elektrotechnik und Informationstechnik und dem Stuttgart Research Centre for Simulation Technology (SRC SimTech) der Universität Stuttgart zur Erlangung der Würde eines Doktors der Naturwissenschaften (Dr. rer. nat.) genehmigte Abhandlung

vorgelegt von

Markus Funk

aus Göppingen

Erstgutachter:	Prof. Dr. Albrecht Schmidt
Zweitgutachter:	Prof. Dr. Gudrun Klinker
Drittgutachter:	Prof. Dr. Shaun Kane
Tag der mündlichen Prüfung:	16. November 2016

Institut für Visualisierung und Interaktive Systeme
der Universität Stuttgart

2016

ABSTRACT

Augmented Reality has been used for providing assistance during manual assembly tasks for more than 20 years. Due to recent improvements in sensor technology, creating context-aware Augmented Reality systems, which can detect interaction accurately, becomes possible. Additionally, the increasing amount of variants of assembled products and being able to manufacture ordered products on demand, leads to an increasing complexity for assembly tasks at industrial assembly workplaces. The resulting need for cognitive support at workplaces and the availability of robust technology enables us to address real problems by using context-aware Augmented Reality to support workers during assembly tasks.

In this thesis, we explore how assistive technology can be used for cognitively supporting workers in manufacturing scenarios. By following a user-centered design process, we identify key requirements for assistive systems for both continuously supporting workers and teaching assembly steps to workers. Thereby, we analyzed three different user groups: inexperienced workers, experienced workers, and workers with cognitive impairments. Based on the identified requirements, we design a general concept for providing cognitive assistance at workplaces which can be applied to multiple scenarios.

For applying the proposed concept, we present four prototypes using a combination of in-situ projection and cameras for providing feedback to workers and to sense the workers' interaction with the workplace. Two of the prototypes address a manual assembly scenario and two prototypes address an order picking scenario. For the manual assembly scenario, we apply the concept to a single workplace and an assembly cell, which connects three single assembly workplaces to each other. For the order picking scenario, we present a cart-mounted prototype using in-situ projection to display picking information directly onto the warehouse. Further, we present a user-mounted prototype, exploring the design-dimension of equipping the worker with technology rather than equipping the environment.

Besides the system contribution of this thesis, we explore the benefits of the created prototypes through studies with inexperienced workers, experienced workers, and cognitively impaired workers. We show that a contour visualization of in-situ feedback is the most suitable for cognitively impaired workers. Further, these contour instructions enable the cognitively impaired workers to perform assembly tasks with a complexity of up to 96 work steps. For inexperienced workers, we show that a combination of haptic and visual error feedback is appropriate to communicate errors that were made during assembly tasks. For creating interactive instructions, we introduce and evaluate a Programming by Demonstration approach. Investigating the long-term use of in-situ instructions at manual assembly workplaces, we show that instructions adapting to the workers' cognitive needs is beneficial, as continuously presenting instructions has a negative impact on the performance of both experienced and inexperienced workers. In the order picking scenario, we show that the cart-mounted in-situ instructions have a great potential as they outperform the paper-baseline. Finally, the user-mounted prototype results in a lower perceived cognitive load.

Over the course of the studies, we recognized the need for a standardized way of evaluating Augmented Reality instructions. To address this issue, we propose the General Assembly Task Model, which provides two standardized baseline tasks and a noise-free way of evaluating Augmented Reality instructions for assembly tasks. Further, based on the experience, we gained from applying our assistive system in real-world assembly scenarios, we identify eight guidelines for designing assistive systems for the workplace.

In conclusion, this thesis provides a basis for understanding how in-situ projection can be used for providing cognitive support at workplaces. It identifies the strengths and weaknesses of in-situ projection for cognitive assistance regarding different user groups. Therefore, the findings of this thesis contribute to the field of using Augmented Reality at the workplace. Overall, this thesis shows that using Augmented Reality for cognitively supporting workers during manual assembly tasks and order picking tasks creates a benefit for the workers when working on cognitively demanding tasks.

ZUSAMMENFASSUNG

Seit mehr als 20 Jahren wird Augmented Reality eingesetzt, um manuelle Montagetätigkeiten zu unterstützen. Durch neue Entwicklungen in der Sensortechnologie ist es möglich, kontextsensitive Augmented-Reality-Systeme zu bauen, die Interaktionen akkurat erkennen können. Zudem führen eine zunehmende Variantenvielfalt und die Möglichkeit, bestellte Produkte erst auf Nachfrage zu produzieren, zu einer zunehmenden Komplexität an Montagearbeitsplätzen. Der daraus entstehende Bedarf für kognitive Unterstützung an Arbeitsplätzen und die Verfügbarkeit von robuster Technologie lässt uns bestehende Probleme lösen, indem wir Arbeitende während Montagearbeiten mithilfe von kontextsensitiver Augmented Reality unterstützen.

In dieser Arbeit erforschen wir, wie Assistenztechnologie eingesetzt werden kann, um Arbeitende in Produktionsszenarien kognitiv zu unterstützen. Mithilfe des User-Centered-Design-Prozess identifizieren wir Schlüsselanforderungen für Assistenzsysteme, die sowohl Arbeitende kontinuierlich unterstützen als auch Arbeitenden Arbeitsschritte beibringen können. Dabei betrachten wir drei verschiedene Benutzergruppen: unerfahrene Arbeitende, erfahrene Arbeitende, und Arbeitende mit kognitiven Behinderungen. Auf Basis der erarbeiteten Schlüsselanforderungen entwerfen wir ein allgemeines Konzept für die Bereitstellung von kognitiver Assistenz an Arbeitsplätzen, welches in verschiedenen Szenarien angewandt werden kann.

Wir präsentieren vier verschiedene Prototypen, in denen das vorgeschlagene Konzept implementiert wurde. Für die Prototypen verwenden wir eine Kombination von In-Situ-Projektion und Kameras, um Arbeitenden Feedback anzuzeigen und die Interaktionen der Arbeitenden am Arbeitsplatz zu erkennen. Zwei der Prototypen zielen auf ein manuelles Montageszenario ab, und zwei weitere Prototypen zielen auf ein Kommissionierszenario ab. Im manuellen Montageszenario wenden wir das Konzept an einem Einzelarbeitsplatz und einer Montagezelle, welche drei Einzelarbeitsplätze miteinander verbindet, an. Im Kommissionierszenario präsentieren wir einen Kommissionierwagen, der mithilfe von In-Situ-Projektion Informationen direkt ins Lager projiziert. Des Weiteren präsentieren wir einen tragbaren Prototypen, der anstatt der Umgebung den Arbeitenden mit Technologie ausstattet.

Ein weiterer Beitrag dieser Arbeit ist die Erforschung der Vorteile der erstellten Prototypen durch Benutzerstudien mit erfahrenen Arbeitenden, unerfahrenen Arbeitenden und Arbeitende mit kognitiver Behinderung. Wir zeigen, dass eine Kontur-Visualisierung von In-Situ-Anleitungen die geeignetste Anleitungsform

für Arbeitende mit kognitiven Behinderungen ist. Des Weiteren befähigen Kontur-basierte Anleitungen Arbeitende mit kognitiver Behinderung, an komplexeren Aufgaben zu arbeiten, welche bis zu 96 Arbeitsschritte beinhalten können. Für unerfahrene Arbeitende zeigen wir, dass sich eine Kombination von haptischem und visuellem Fehlerfeedback bewährt hat. Wir stellen einen Ansatz vor, der eine Programmierung von interaktiven Anleitungen durch Demonstration zulässt, und evaluieren ihn. Bezüglich der Langzeitwirkung von In-Situ-Anleitungen an manuellen Montagearbeitsplätzen zeigen wir, dass Anleitungen, die sich den kognitiven Bedürfnissen der Arbeitenden anpassen, geeignet sind, da ein kontinuierliches Präsentieren von Anleitungen einen negativen Einfluss auf die Arbeitsgeschwindigkeit von erfahrenen Arbeitenden sowohl als auch unerfahrenen Arbeitenden hat. Für das Szenario der Kommissionierung zeigen wir, dass die In-Situ-Anleitungen des Kommissionierwagens ein großes Potenzial haben, da sie zu einer schnelleren Arbeitsgeschwindigkeit führen als traditionelle Papieranleitungen. Schlussendlich führt der tragbare Prototyp zu einer subjektiv niedrigeren kognitiven Last.

Während der Durchführung der Studien haben wir den Bedarf einer standardisierten Evaluierungsmethode von Augmented-Reality-Anleitungen erkannt. Deshalb schlagen wir das General Assembly Task Model vor, welches zwei standardisierte Grundaufgaben und eine Methode zur störungsfreien Analyse von Augmented-Reality-Anleitungen für Montagearbeiten bereitstellt. Des Weiteren stellen wir auf Basis unserer Erfahrungen, die wir durch die Anwendung unseres Assistenzsystems in Montageszenarien gemacht haben, acht Richtlinien für das Gestalten von Montageassistenzsystemen vor.

Zusammenfassend bietet diese Arbeit eine Basis für das Verständnis der Benutzung von In-Situ-Projektion zur Bereitstellung von kognitiver Montageassistenz. Diese Arbeit identifiziert die Stärken und Schwächen von In-Situ-Projektion für die kognitive Unterstützung verschiedener Benutzergruppen. Folglich tragen die Resultate dieser Arbeit zum Feld der Benutzung von Augmented Reality an Arbeitsplätzen bei. Insgesamt zeigt diese Arbeit, dass die Benutzung von Augmented Reality für die kognitive Unterstützung von Arbeitenden während kognitiv anspruchsvoller manueller Montagetätigkeiten und Kommissioniertätigkeiten zu einer schnelleren Arbeitsgeschwindigkeit führt.

PREFACE

This thesis contains work created from 2013 till 2016 at the University of Stuttgart. Since the development of an assistive system requires different types of expertise from different disciplines, this thesis has been done in close collaboration with experts from the University of Stuttgart, projects partners within the motionEAP project, and external collaborators. These collaborations resulted in publications which are a core part of this thesis. The contributing authors (i.e., co-authors of papers) are clearly stated at the beginning of each chapter together with the reference to the publication if applicable. To keep the consistency throughout the thesis and to emphasize these collaborations, I use the term *"we"* instead of *"I"* when referring to myself.

Acknowledgments

Over the past 4 years, I had the pleasure of working together with a number of amazing researchers and persons which inspired me and helped me a lot with the research and work that led to this thesis. First and foremost, I would like to thank my supervisor **Albrecht Schmidt** who inspired me with his passion for the field of Human-Computer Interaction. His creativity, ideas, and guidance helped me with the creation of the papers that are included in this thesis. I am still impressed how he is able to give such in-depth feedback with a travel schedule as busy as his. Further, I would like to thank my committee **Gudrun Klinker**, **Shaun Kane**, and **Stefan Wagner** for the great discussions and feedback.

On the journey that led to this thesis, there is **Lars Erik Holmquist** that I would like to thank especially for taking me as an intern at Yahoo! Labs and for encouraging me to start my PhD. I learned a lot during the time I was working with him in California.

I further had the pleasure to work in the best research group that one could ever imagine - the HCILab group in Stuttgart, which gathers amazing persons and researchers. At first, I would like to particularly thank **Alireza Sahami Shirazi** for his guidance and teaching me how to work scientifically when I was still a young PhD student. Without him, I certainly would not have been able to *"Schreibe die Paper"*. Further I would like to thank a number of researchers and persons that helped and influenced me a lot with their great personality and their research experience: **Stefan Schneegass** (for being there for me whenever I needed help and for being an awesome office and conference-room mate), **Florian Alt** (for being an awesome office mate although he supports the wrong soccer team), **Niels Henze** (for showing me that text entry studies and Fitts' law experiments can be very interesting). Further, I am very grateful to **Sven Mayer** for being an amazing friend and for his extraordinary support during tough times in the project. Moreover, I would like to thank **Anja Mebus** for always having an open ear for me whenever I needed something.

I would also like to thank the rest of the awesome HCILab group in Stuttgart. You guys are making every day a great day! Especially I would like to thank **Pascal Knierim** (for being an awesome friend and co-pilot), **Thomas Kubitza** (for unforgettable NYC stories and showing me how to play foosball), **Tilman Dingler** (for showing me great whiskey and countless cheese experiences), **Katrin Wolf** (for showing me the Merlot wine), **Bastian Pfleging** (for helping me with complicated tax returns), **Lars Lischke**, **Yomna Abdelrahman** (for amazing Egyptian food), **Alexandra Voit**, **Mauro Avila**, **Passant El.Agroudy**, **Miriam**

Greis, **Francisco Kiss**, **Tonja Machulla**, **Lewis Chuang** (for properly teaching me statistics), **Pawel Wozniak**, **Dominik Weber**, **Valentin Schwind**, **Patrick P. Bader** (the P. is for president) and **Huy Viet Le**.

I am very grateful to being able to work in the amazing #teammotionEAP of the University of Stuttgart over the last 4 years: I would like to thank **Thomas Kosch** (for countless fun Fridays and drone challenges - thank you for dealing with my craziness), **Romina Kettner** (for awesome team motionEAP trips), **Hauke Behrendt**, **Michael Matheis**, **Mai ElKomy**, and **Oliver Korn**. Further I would like to thank all the students which I had the pleasure of working with during the motionEAP project.

Another very important part of the last 4 years was my research visit to the MIT Media Lab. Therefore, I would like to deeply thank **Scott W. Greenwald** (for being an awesome house mate and friend) and **Pattie Maes**.

Further, I would like to thank **Nicolaj Gruzdov** and **Jasmin Trautner** for being there for me in the best and worst times! Without my two best friends I would not be the person I am today. I am also very grateful for countless discussions with **Henner Schmidt**. Thank you for supporting me with your cybernetics knowledge.

I want to thank my family for their unconditional support: My parents **Sabine Funk** and **Wolfgang Funk** for raising me to be the person I am today, for making it possible for me to explore technology, and for supporting me in getting the education I wanted. Further, I would like to thank my sister **Claudia Funk** for being there for me and supporting me. I sincerely apologize for being away that often.

Last but not least, I would like to thank **Marzieh Zonouzi**, for being an amazing person, being there for me, and supporting me through all the good and bad times that occurred when finishing this thesis.

TABLE OF CONTENTS

List of Figures

List of Tables

List of Acronyms

GUI	graphical user interface
NASA-TLX	NASA-Task Load Index
ER	error rate
TCT	task completion time
HMD	head-mounted display
AR	Augmented Reality
VR	Virtual Reality
PI	Performance Index
TPB	time per brick
EPB	errors per brick
IAR	Industrial Augmented Reality
SAR	Spatial Augmented Reality
SURF	Speeded Up Robust Feature
CMD	cart-mounted display
PbVi	Pick-by-Vision
PbVo	Pick-by-Voice
PbP	Pick-by-Paper
PbL	Pick-by-Light
FoV	Field of View
GATM	General Assembly Task Model
PbD	Programming by Demonstration
CWD	chest-worn display

To avoid confusions with the naming of the Microsoft Kinect sensors, throughout this thesis, we refer to the first Microsoft Kinect for Windows as **Kinect_v1** and to the newer Microsoft Kinect "One" as **Kinect_v2**.

I

Introduction and Motivation

Chapter 1

Introduction

Over the last twenty-five years, computers evolved from being stationary and huge systems that were accessed by using terminals to being integrated into nearly every aspect of people's lives. Although personal computers still exist, computer systems truly became ubiquitous [154] as they now support nearly every aspect of people's lives without them noticing anymore. Computers are used for providing access to information, facilitating planning tasks, and for controlling the users' homes according to their habits and needs.

Using computers also changed the process of manufacturing products by optimizing workflows, facilitating the organization of a supply chain, and being able to use computers for machine maintenance tasks and training workers. For example, Virtual Reality (VR) learning systems are enabling risk-free learning in safety-critical tasks [15]. Moreover, Industrial Augmented Reality (IAR) is now present in every aspect of a manufactured product's life-cycle (cf. [117]). Experiencing a designed product can be done immediately [42], industrial robots cooperating with human workers can be programmed using Augmented Reality (AR) debugging approaches [33], and maintaining existing machines and products can be supported directly on site [160].

Also ordering products is nowadays supported by a computer system. A great benefit for customers is that they can configure a product that they are ordering according to their individual needs. Considering the manufacturers, enabling the customers to customize nearly every aspect of a product leads to a huge diversity of variants. As storing every possible variant of a product would increase the storage costs, the manufacturers nowadays tend to produce an ordered product *on-demand*, just when the customer ordered it. This leads to a shift in how products are being produced, as in the assembly process every assembled product will be different from the previously assembled product. This is also known as producing in *lot size one*. The challenge in manually assembling products in *lot size one* is the increased cognitive effort for assembly workers, as the workers always have to be concentrated to assemble a product exactly as specified in the current order.

Teaching new or inexperienced workers how to assemble a manufactured product is an important topic in the manual manufacturing industry. Especially when producing many variants of a product, processes become more complex and are harder to learn. In some companies video tutorials are used to teach workflows, but traditionally a more experienced colleague is asked to teach a workflow to an inexperienced worker [114]. In the recent years, computer-aided assistive systems have been proposed for supporting manual assembly workers in cognitively demanding assembly tasks. Assistive systems providing interactive AR instructions [32] have been suggested to assist workers during assembly tasks. For order picking tasks, Pick-by-Light (PbL) systems visually show a worker, where the next part has to be picked from. Also head-mounted displays (HMDs) can show the position of the next part to a worker and where the part has to be assembled [149].

One of the most important areas of application of such assistive systems is the support and inclusion of impaired workers into the working life [83]. Through continuously providing instructions, impaired workers can work on more complex products [139] which enables a better integration. This additionally fosters the inclusion of impaired workers into a company's daily business and has the potential to increase the productivity in sheltered work organizations. The requirements for good work and the potentials of assistive systems to contribute to an inclusion impaired workers are outlined by Behrendt et al. [19].

1.1 Research Questions

Assistive systems for the workplace using in-situ projection for providing instructions at workplaces have the potential for becoming ubiquitously available. Therefore, research is required to identify potentials and limitations of assistive technology using in-situ projection.

In this thesis, the evaluation of assistive systems using in-situ feedback is designed to follow a bottom-up approach. By first identifying the optimal design of the in-situ instructions, we move towards creating instructions and finding metrics for evaluating them. The research questions of this thesis also follow this bottom-up approach (see Table 1.1).

We start with exploring the design of visual in-situ instructions. Considering that one application area for using assistive systems at the workplace is integrating cognitively impaired workers into the work life, the ways of providing feedback have to be specially tailored to their requirements. Thus, the different possibilities to visualize assembly instructions using in-situ projection have to be explored for cognitively impaired workers (RQ1). Further, using in-situ instructions might have a huge benefit for integrating cognitively impaired workers into the work life. Hence, the potentials of using assistive systems for supporting cognitively impaired workers during assembly tasks need to be explored (RQ2). Assistive systems have the potential to not only convey instructions at the workplace where they are needed. Due to being context-aware, they can also react to potentially made errors and prevent them from happening. Considering that different types of error feedback can have different implications, it is important to explore the possibilities of multi-modal error feedback at manual assembly workplaces (RQ3).

Once we have evaluated how in-situ instructions for assistive systems should be designed, we need to evaluate how these in-situ instructions can be created in a time-saving and cost-reducing way. Therefore, different alternatives for creating interactive assembly instructions need to be compared in a lab environment and on a real product (RQ4).

For exploring the external validity, we are interested in learning, how long in-situ instructions at the manual assembly workplace are creating a benefit for the workers in real production scenarios. Therefore, research needs to be conducted to evaluate the long-term effects of in-situ instructions (RQ5). Manual assembly is only one part of the process of producing products. Therefore, we are interested in transferring the knowledge to other areas of the production process. An error-prone task in the process of producing products that can benefit from in-situ instructions in is order picking. Hence, we transfer the concept of in-situ instructions to order picking scenarios (RQ6). When designing a new way of presenting an instruction, the instruction is evaluated with the product that the instruction was designed for. However, comparing different types of instructions to each other is cumbersome, because instructions are usually only evaluated with a special use case. Thus there is a need for creating a benchmark task and creating an evaluation process for making instructions comparable to each other (RQ7).

Research Question	No.	Chapter
What are suitable in-situ visualizations for assembly instructions?	(RQ1)	Chapter 8
How are in-situ instructions perceived by cognitively impaired workers?	(RQ2)	Chapter 8
Which modality can communicate errors best to workers?	(RQ3)	Chapter 9
How can in-situ instructions for assistive systems be created ?	(RQ4)	Chapter 10
What are the long-term effects of using in-situ instructions?	(RQ5)	Chapter 11
How can in-situ instructions be used for order picking tasks?	(RQ6)	Chapter 12
How can instructions for workplaces be evaluated?	(RQ7)	Chapter 13

Table 1.1: An overview of the research questions addressed in this thesis.

1.2 Methodology

Using AR to provide instructions for complex work tasks was first suggested by Caudell and Mizell [32] in 1992. Since then, a body of research projects focused on combining AR technologies and the manufacturing domain. However, these prototypes mostly remained at a proof-of-concept stage. In 2004, Navab [117] described requirements for *"killer applications"* in IAR that have the opportunity to make their way out of the research lab into becoming a product that is usable in every day work. Since then, a few assistive systems using in-situ projection made it to the market. However, research in providing assembly instructions for a manual assembly workplace was only conducted in lab studies and did not use the full potential of in-situ projection.

Designing an assistive system using IAR for real applications scenarios requires a multi-disciplinary team including computer science, mechanical engineering, psychology, and philosophy. For closing the gap between these disciplines and finding a common basis for discussion, we aimed to create an assistive system for testing different aspects of AR rapidly. Following a user-centered design process with multiple iterations, it was our aim to build a functional prototype quickly. This bottom-up approach enabled us to conduct fundamental user studies at the beginning of the project (e.g. Chapter 8). With the help of the gained feedback from target users and continuously improving the prototype in multiple iterations, we created a fully functioning assistive system that is robust enough to conduct a long-term study in a real manufacturing scenario.

1.2.1 Prototypes

The research prototypes presented in this thesis are a result of a collaborations with my colleagues, the work of student assistants, and projects of undergraduate students. Especially valuable in designing the prototypes was the user feedback that we received from our industrial partners when testing the prototypes. Taking their feedback into account, we could design the prototypes in a way that they were especially tailored to the requirements of the workers. Further, the prototypes were designed in a way to be robust and to be deployed in real manufacturing factories. Therefore, the prototypes were developed after high quality standards and had to be tested thoroughly before we deployed them in the factories.

1.2.2 Evaluation

As the goal of this thesis is to built an assistive system that is capable of supporting a wide range of user groups (from cognitively impaired workers to experienced workers), we followed a user-centered design[1] process with multiple iterations that first identifies the requirements, creates a design, builds a prototype and evaluates it. Through qualitative feedback sessions with experts from our industrial partners, experienced workers, and cognitively impaired workers, we improve our assistive system with each iteration to fit the requirements of the user groups.

For the evaluation part of the user-centered design process, we conducted two types of studies: controlled lab studies and field studies in real production environments. However, as a field study is very costly and the assistive system has to work very robustly for conducting a field study, we considered performing the lab studies first. In the lab studies, we analyze single aspects of an assistive system and find the best design of the feedback before using the system in a real production environment (cf. Chapter 8 and Chapter 9). Both in controlled lab studies and field studies, we perform a quantitative analysis to evaluate the effects of an assistive system on production parameters, e.g. the time to produce a part, the errors that are made, or the perceived workload. Further, we collect qualitative feedback for improving the usability of the assistive system.

---

[1] ISO 9241-210 - `http://www.iso.org/iso/catalogue_detail.htm?csnumber=52075` - (last access 5th Oct. 2016)

1.2.3 Ethics

The research that was conducted in the scope of this thesis is part of the project motionEAP that was funded by the German Federal Ministry for Economic Affairs and Energy. Before this project was approved, it went through an ethical approval process. Additionally, for every user study that was conducted in factories of industrial partners as a part of this thesis, we went through an ethical approval process that was conducted by the works council of the factories where the studies were conducted. As a part of the motionEAP project, Hauke Behrendt [19] is presenting ethical considerations arguing why assistive systems can improve the work quality and foster inclusion of impaired workers in the work life.

1.3 Research Context

The research that lead to this thesis was conducted over the course of three years at the *University of Stuttgart* (group for Human-Computer Interaction). During this time, the research was inspired by collaborations, publications, and discussions with many experts from different areas.

motionEAP
The first major part of the research reported in this thesis was conducted in the project motionEAP. motionEAP is a project funded by the German Federal Ministry for Economic Affairs and Energy. A comprehensive overview about the project goals, publications, and events that resulted in the project can be found on the project website[2]. The motionEAP consortium consists of outstanding researchers from the fields of computer science, philosophy, pedagogy, and mechanical engineering. Of particular importance were the collaborations with Oliver Korn, Hauke Behrendt, Liane and Andreas Bächler, and Thomas Heidenreich, which led to a number of publications (e.g. [19, 45, 50, 51, 93, 94, 95, 96]).

University of Stuttgart
The second major part of the research reported in this thesis was conducted together with colleagues from the University of Stuttgart. Combining the technical knowledge and scientific expertise of the group with my research interests, resulted in a number of publications that are of great importance for this thesis. The collaboration with Stefan Schneegass, Niels Henze, Alireza Sahami Shirazi, Tilman Dingler, Katrin Wolf, Bastian Pfleging, Yomna

[2] motionEAP website: `http://www.motioneap.de` - (last access 5th Oct. 2016)

Abdelrahman, Pascal Knierim, Sven Mayer, Lars Lischke, Thomas Kosch, and Albrecht Schmidt led to publications within the scope of this thesis (e.g. [48, 53, 57, 58, 61, 64]) and further publications beyond the scope of this thesis (e.g. [1, 8, 38, 46, 47, 56, 60, 62, 63, 100, 101, 144, 162]).

Of particular success was the collaboration with Sven Mayer, which resulted in a publication that was nominated for the best paper award at the ASSETS'15 conference [59].

External Collaborations
Further research beyond the scope of this thesis was conducted with external colleagues. The collaboration with Sebastian Büttner from the Ostwestfalen-Lippe University of Applied Sciences led to a publication [29]. A further collaboration with Scott Greenwald from the MIT Media Lab led to a publication comparing Augmented Reality and Virtual Reality learning approaches [68].

1.4 Outline

This thesis consists of 15 chapters that are grouped into five parts. The *Background* part of this thesis provides an overview of the motivation behind this research, provides an overview about commercial systems, and introduces related work that was previously published in this area. The *System* part describes the concepts and implementation behind the assistive system that was developed as main instrument in this thesis. In the *Evaluation* part, the previously described system is evaluated considering designing, creating, and evaluating context-aware in-situ instructions. The *Results* part contains a summary of the contributions and findings and provides ideas for future work. A graphical outline of this thesis is depicted in Figure 1.1.

Part II: Background

Chapter 2 - Background: This chapter describes the motivation for this work, introduces terms and definitions in the area of assistive systems for the workplace, and briefly introduces the experimental approaches that were used throughout this thesis. Finally, this chapter provides an overview about commercially available assistive systems that are used for training new workers for an assembly task, or to continuously provide quality support.

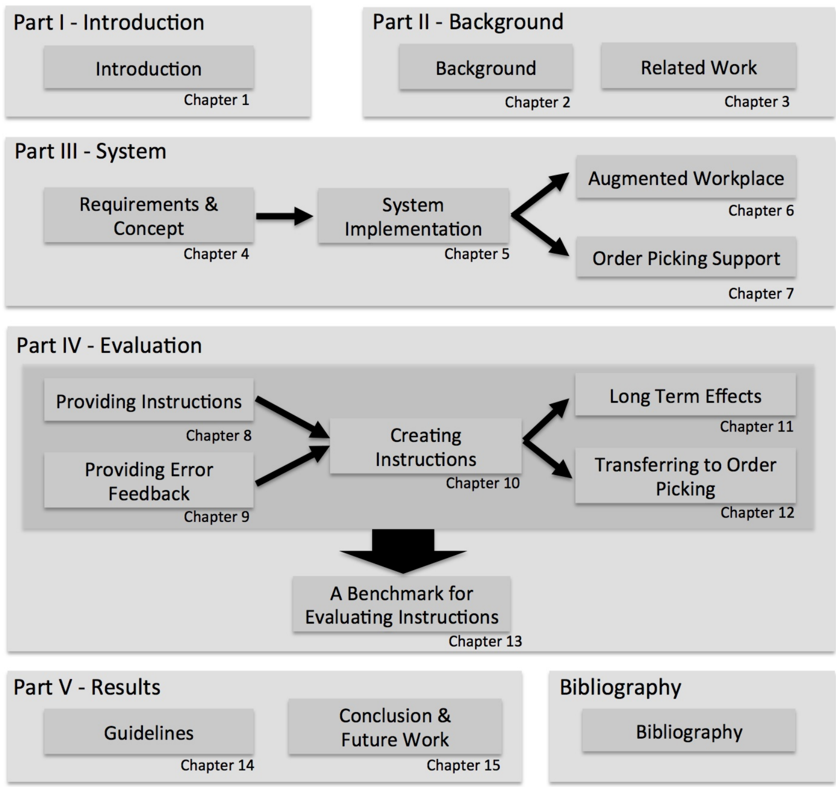


Figure 1.1: A graphical representation of the thesis outline.

Chapter 3 - Related Work: For providing an overview about related work, we summarize research projects in the areas of Augmented Reality, sensing interaction, assistive systems at the workplace, instruction giving in assembly, and order picking support systems.

Part III: System

Chapter 4 - Requirements and Concept: As an outcome of interviews with experts and users, we outline requirements for a assistive systems for supporting workers at the workplace. After defining the interaction logic of the assistive system, we describe use cases and list functional technical requirements that an assistive system has to fulfill. According to the described requirements, we suggest a general architecture for an assistive system. As the interaction with an assistive system should be as natural as possible, we describe envisioned interaction concepts: e.g. creating instructions through Programming by Demonstration (PbD), interacting with the system implicitly using work step detection, and using user-defined tangible objects as input for digital functions.

Chapter 5 - System Implementation: This chapter introduces a software architecture that fulfills the requirements that were outlined in Chapter 4. It was designed explicitly to be independent from the underlying technology to be able to transfer the software architecture to different use cases. The architecture comprises a workflow logic that is responsible for mapping the visual feedback to the corresponding work steps. The *Scenes and Triggers* concept further enables an adaptive feedback that adjusts the provided feedback according to the user's performance. In the second part of this chapter, we introduce technology dependent algorithms for detecting activity at the workplace. Through a pick detection, assembly detection, and object recognition, the assistive system can detect most of the actions that are performed at the workplace. Finally, the concept is deployed in four prototypes: the manual assembly workplace, three interconnected workplaces - the assembly cell, and in two order picking systems - OrderPickAR and HelmetPickAR.

Chapter 6 - An Augmented Workplace: As one major contribution of this thesis is the assistive system for providing Augmented Reality instructions at the manual assembly workplace, the built system is described in detail in this chapter. First, the hardware setup of a single assembly workplace is described. Afterwards, the two user roles engineer and worker are introduced and their tasks are described. According to these tasks, the user interface and the interaction design is described for both user roles. Thereby, the natural user interface using activity recognition and the user interface using a graphical user interface on a monitor are described. Further, the chapter explains the concept of using adaptive feedback and introduces the three levels beginner, advanced, and expert. It is explained according to which criteria the system decides to switch between the adaptivity levels. Finally, the chapter introduces a vision for a product which is providing Augmented Reality instructions at a single manual assembly workplace.

Chapter 7 - An Order Picking Support System: In this chapter, we introduce two prototypes of an assistive system for order picking, which use the concepts and software architecture that is described in the previous chapters. The two systems address two possible alternatives of the technology-placement design dimension of assistive technology: technology is placed in the environment and technology is user-worn. The first prototype *OrderPickAR* augments an order picking cart with camera-projector pairs. Thereby, context-sensitive picking instructions can be shown to the user. In the second prototype, the *HelmetPickAR*, the worker is wearing an interactive helmet containing a camera-projector pair. Both prototypes are introduced and described in this chapter.

Part IV: Evaluation

Chapter 8 - Evaluation of Feedback Mechanisms for Workplaces: The key requirement for supporting workers at the workplace is the quality of the instruction. Hence, as a first step, we are interested in finding the most suitable visualization of in-situ instructions. Through a user study with cognitively impaired workers, a contour visualization is compared to a pictorial visualization, a video visualization, and a baseline without in-situ instructions. The results reveal that a contour visualization was perceived best by the participants. As a second step, we are interested in evaluating the effect of the contour-based in-situ instructions on cognitively impaired workers. Through a user study, we evaluate the potentials of continuously providing in-situ instructions at the workplace for impaired workers. The results show that using in-situ instructions cognitively impaired workers are capable of performing more complex tasks at a manual assembly workplace.

Chapter 9 - Evaluation of Error Feedback for Workplaces: After learning about how to present work instructions, this chapter focuses on how to optimally present feedback that communicates the worker that an error was made. Traditionally the error feedback is presented visually, but considering privacy and distraction, other modalities might be considered. Through a user study, we compare haptic, auditory, and visual notifications as error feedback at manual assembly workplaces. The results show that using haptic feedback retains the workers privacy.

Chapter 10 - Evaluation of Instruction Creation Methods: Creating context-aware in-situ instructions usually requires programming effort, which is time-consuming. In this chapter, we introduce a mechanism to program a workflow for an assistive system by demonstrating it once. In a first study, we compare instructions that were created using this Programming by Demonstration concept to traditional assembly videos. In a second study with experienced workers, we measure the time it requires to create context-aware in-situ instructions compared to traditional assembly videos and a graphical editor. Lastly, we compare the created instructions to each other by instructing inexperienced workers to assemble real products.

Chapter 11 - Evaluation of Long-Term Impact: One of the key use cases for assistive systems is using them continuously as a quality support system, which only interferes if workers are about to make an error. Therefore, we are interested in evaluating the effects of using an assistive system over a long period of time. In a field study with experienced workers and inexperienced workers, we provide in-situ instructions in a real assembly scenario for three workdays. The results indicate that workers are able to successfully learn the assembly steps using in-situ projection. However once the workers know the assembly steps, the in-situ instructions are slowing the workers down.

Chapter 12 - Comparing Modalities for Order Picking: When transferring the concept of using in-situ projection to cognitively support workers for order picking, the design dimension of where to put the technology plays an important role. In this chapter, we evaluate our body-worn HelmetPickAR prototype and the cart-carried OrderPickAR prototype. The results show that using HelmetPickAR decreases the perceived workload, however the placing time increases for non-complex tasks. In contrast, the cart-carried OrderPickAR outperforms traditional approaches e.g. Pick-by-Paper, Pick-by-Voice, and Pick-by-Vision considering task completion time and is well perceived by the workers.

Chapter 13 - Towards a Benchmark for Instruction Giving: Comparing interactive instructions to each other is cumbersome as every instruction giving system uses its own task for the evaluation. In this chapter, we propose using two reference tasks for evaluating instruction giving systems. Further, we provide and evaluate a formula that groups the task-completion time into task-dependent and task-independent measures. In a user study, we use the proposed evaluation technique by comparing our in-situ projection approach to a tablet-based approach, to instructions shown on an HMD, and to a paper-baseline. The results show that using the in-situ projection and the paper-baseline outperform the HMD and tablet instructions.

Part V: Results

Chapter 14 - Guidelines for Assistive Systems: Based on the experiences in designing assistive systems and conducting a number of user studies with different user groups, we provide general guidelines and recommendations for designing assistive systems.

Chapter 15 - Conclusion and Future Work: A summary of the contributions of this thesis is provided in the conclusion. Finally, an overview about interesting research topics that are connected to this thesis are provided.

II

Background

Chapter 2

Background

Using Augmented Reality (AR) feedback for supporting manual assembly tasks addresses real problems that can be found in many manufacturing companies nowadays. In this chapter, we provide an overview about the motivation for this work, provide an overview about the different user groups that are involved in manual assembly, and introduce terms and definitions that are used in this thesis. Further, we describe our general approach for identifying and addressing research gaps in this area and finally provide an overview about commercially available assistive systems.

2.1 Motivation

In the 2011 world report on disabilities [120], the World Health Organization (WHO) describes which conditions are seen as a disability or impairment. They state that *"almost everyone will be temporarily or permanently impaired at some point in life"* [120]. The report shows that disability increases with an increasing age, however, the aging persons will not consider themselves as being disabled as they consider the disability as a part of the aging process. In the WHO report on disabilities, the authors used a similar method as the Performance Index, which we use in most of our studies to quantify the degree of impairment. For their statistics the WHO uses a threshold of 40, which equals to a Performance Index of 60%. According to their report [120] in which they derive from data collected from 59 countries, 8.9% of the age group between 18 and 49 are considered to have a disability. In the age group between 50 and 59 the number of persons with disabilities is 20.6% across the population.

The fact that especially older persons are more likely to having a disability, leads to an increasing amount of workers with disabilities that are employed in Germany, as the work society in Germany is experiencing a drastic change in its average age. Due to the increase in general life expectancy and the decrease in birthrate [127], the society as a whole is becoming older. This results in a change of the number of persons that are in a working age. According to the German Federal Office of Statistics [28], the number of persons that are in a working age will decrease. If the immigration rate will retain at a level up to 200.000 immigrants per year, the number of persons in a working age will drop by 23% until the year 2060 compared to the year 2013. However, if the immigration rate will also drop to 100.000 immigrants per year, the number of persons in a working age will drop by up to 30% until the year of 2060 compared to the year 2013. This drop in number of persons in a working age results in two changes for the working society. First, the working society will have to acquire a broader set of skills due to the reduced number of persons that are available for the jobs. Second, the average age in the working society will increase. Therefore, the need for cognitive assistance in work tasks might increase.

The world report on disabilities [120] mentions that there are inequalities considering employment when comparing persons with disabilities and persons without disabilities. According to this report, the persons with disabilities are less likely to be employed and are earning less money when employed. Therefore, there is still work to be done to include persons with disabilities in the work life and to enable persons with disabilities to get equal chances and to get paid equally. Also the UN convention on the rights of persons with disabilities [83] especially outlines this need for equal chances. Further the convention describes steps for including persons with disabilities in their article 27. It has been argued that assistive systems can foster inclusion of persons with disabilities into the world of employment [19, 83]. Therefore creating cognitive assistance technology for the workplace might lead to an increased inclusion and equal chances for workers with disabilities.

Another way of using cognitive assistance technology at the workplace is to teach inexperienced workers how assembly steps are being performed. As especially in Germany, we are experiencing a trend that companies hire more temporary workers, the number of inexperienced workers is increasing. According to the German Federal Employment Agency [65], there were 961.000 temporary workers employed in Germany in 2015. This is an increase of 240% compared to 282.000 temporary workers in 2003. These workers sometimes have a general training in manual manufacturing, however they need to be trained to assemble a specific product first.

Another important factor that has to be considered when thinking about the cognitive effort that is required in manual assembly scenarios is that the number of variants that are produced in the manufacturing industry are constantly increasing. Franke [44] introduces a summary of reasons for the increasing number of produced variants, e.g. political changes in the market, new areas of application, or retaining support for produced parts. Another reason might be that the storage cost for assembly parts is increasing and therefore the parts are rather produced on-demand than produced to store them in a warehouse until they are needed. Interestingly, this trend of producing many variants of a product and customizing a product according to a customer's needs is recently leading big car manufacturing companies to employ more human workers again[3]. As human workers can deal with these many variants better than automatic robots, the companies tend to use robots only as a tool to assist humans during the assembly tasks. Hereby, the main problem of industrial robots is that they have to be programmed explicitly for each variant. Also switching between the programs for the robots is an issue, as not only the actions of the robot have to fit the variant, but also the physical arrangement of the workplace has to be changed accordingly. Therefore, a huge number of human workers are used in these variant-rich assembly tasks. However, producing many variants is a complex task that can lead to an increased cognitive load for the workers at the manual assembly workplace, especially as every assembled product might be different from the previously assembled product. Thus assisting these workers by reducing their cognitive load is a task that assistive systems for the workplace should address.

2.2 User groups in assembly environments

When analyzing the user groups that are involved in working on manual assembly tasks in the industry, we can split the population in three user groups:

Experienced workers, are skilled manual workers trained in manual assembly and who have experience with the manufactured product. Usually they were

[3] Bloomberg Business - `http://www.bloomberg.com/news/articles/2016-02-25/why-mercedes-is-halting-robots-reign-on-the-production-line` - last access 5th Oct. 2016

gaining this experience by assembling the product for several years. Experienced workers know every detail about assembly steps, know work steps which are error-prone, and know how to avoid these errors best.

Inexperienced workers, are manual workers trained in manual assembly but who have no experience with the manufactured product as they are assembling the product for the first time. Due to rapidly changing requirements in assembly more and more manual workers are used for tasks in which they are not experienced. According to a survey of the German Socio-Economic Panel [74], only 5.72% of the workers that were employed from 1999 to 2011 had a low qualification. Especially nowadays where a mismatch in skills and jobs lead to a skills shortage[4], the number of inexperienced workers will increase. Also the workers with a temporary contract are considered inexperienced workers. As contracts of temporary workers are usually limited to six weeks, training new temporary workers is required very often. Historically, inexperienced workers are trained by experienced workers before they are able to assemble a product themselves.

Workers with impairments, are skilled manual workers that received a training in manual assembly, but who have special needs due to their cognitive or physical impairment. Following the UN convention on the rights of persons with disabilities [83], more and more companies are integrating workers with impairments by offering special work places for conducting easier assembly tasks. In 2013, approximately 7 million workers with impairments were employed in the United States [41]. As supporting workers with impairments requires special training for instructors, integrating workers with impairments is often outsourced to sheltered work organizations [99] where trained socio-educational instructors are continuously supporting the workers with impairments. Here we see a great benefit of assistive systems as they might have a great impact on supporting workers with impairments continuously.

2.3 Terms and Definitions

In this section, we introduce terms and definitions that are used throughout this thesis and that are used in industrial environments. Although there are previous definitions of these terms, we introduce them exactly the way we used them with our project partners in industrial manufacturing contexts.

[4] BBC news - `http://www.bbc.com/news/business-34297368` - last access 5th Oct. 2016

Lot size
Lot size is a number that indicates how many items of one product are produced until the assembly line or the workplace produces an other product. For example, producing a product in lot size one, means that only one item of the product is produced at one workplace. The next produced item is a different product than the one produced before.

Manufacture on demand
For reducing storage costs and because customers are nowadays able to highly customize ordered products, many products are only manufactured when they were ordered by the customer in their final version. Some years ago being able to manufacture on demand was very cumbersome and logistically not feasible. Therefore, the ordered products were produced in batches and were stored in a warehouse until a customer ordered them.

Industry 4.0
The term industry 4.0 means using information technology in general for improving industrial manufacturing processes. Thereby the improvement can be made in any state or stage of the process, e.g. using information technology to improve the human-computer interface for a machine with digital components, or using sensors and actuators to be aware of the manufacturing status at all times. The term itself refers to the 4th industrial revolution, which is using information technology in industrial manufacturing processes. The first revolution was using water and steam for producing, the second revolution was building assembly lines for production, and the third industrial revolution was using automated robots for the production lines [88]. An overview of all German projects that are addressing industry 4.0 issues are listed online[5].

Augmented Reality
Although there are a lot of definitions for Augmented Reality (AR), in this thesis we define it as follows: Augmented Reality is the combination of digital information with real world scenarios or physical objects according to real world circumstances. Thereby, the information can have any modality and is not limited to visual information.

Spatial Augmented Reality
In this thesis Spatial Augmented Reality (SAR) is used as visual Augmented Reality information that is registered at a fixed point in the physical space. Thereby the used technology for augmenting the reality spatially is not specifically defined, as SAR can be achieved e.g. using HMDs, projectors, or hand-held screens. With

[5] Platform Industry 4.0 - `http://www.plattform-i40.de` - last access 5th Oct. 2016

using SAR, the position of the visual information in the space can either be a fixed X/Y/Z coordinate or the visual information can be spatially attached to a movable object.

Industrial Augmented Reality
In the context of this thesis, we are using the term Industrial Augmented Reality as using AR for industrial processes. Thereby, the information does not have to be spatially registered. According to Navab [117], IAR applications can be assigned to the following categories: design, commissioning, manufacturing, quality control, monitoring and control, service and maintenance.

Assistive System
As an assistive system, we define an interactive system using a technology to give instructions or feedback to a worker while performing work tasks. Thereby the modality that is used for presenting the instructions or the technology that is used to present them is not specified. An assistive system can use one or many modalities or technologies to present instructions or feedback during work tasks. Further, an assistive system can be, but does not have to be, context-aware, i.e. reacting to a user's actions or tasks. In the literature, assistive systems are also referred to as assistance systems, interactive instructions, or instruction systems.

2.4 Approach

The research approach used in this thesis is a bottom-up approach with three iterations (phases). Each iteration follows the user-centered design process consisting of the following steps: First, requirements were collected through interviews with different user groups and experts from the industry according to which a design was created. Secondly, we implemented the design in our assistive system to rapidly being able to use the aspects of the design which needed to be studied. Lastly, from the requirements that were identified in the first step, we formulate hypotheses that we want to accept or reject. To being able to scientifically evaluate the correctness of the formulated hypotheses, we conduct empirical user studies using the prototypical implementation of the design and testing different aspects of the design with users.

Considering the results of the user studies of the previous phase, we afterwards repeat the user-centered design process and address a new aspect of the assistive system. Within the scope of this thesis, we considered dividing the research into three phases. A graphical representation of the approach is depicted in Figure 2.1.

Phase 1: Presenting instructions and communicating errors
In the first phase, we address how instructions are presented using an assistive system for the workplace and how to best communicate when an error was made. To achieve this, we started the exploration part of phase 1 with doing a literature review in the areas of AR, assistive systems for the workplace, and general instruction psychology. Further, we conducted interviews with experts from the industry to identify relevant state-of-the-art approaches. As the target users of our assistive system are both experienced workers and cognitively impaired workers, we talked to workers and supervisors working in both domains. Afterwards, we created a design for both presenting instructions using in-situ projection and for error feedback and created a software architecture, which supports experimenting with different instructions and error feedback modalities. In the examination part of phase 1, we implemented the envisioned design and conducted three user studies. The results of the user studies conducted in phase 1 are reported in Chapter 8 and Chapter 9.

Phase 2: Creating instructions The second phase addresses how instructions that are presented by assistive systems for the workplace can be created with little effort and how accurately the created instructions can convey knowledge about assembly steps to inexperienced workers. In the exploration part of phase 2, we conduct a market analysis and a literature overview to learn about the state-of-the-art instructions and how they are created when introducing new products. Further, we created a design to transfer a Programming by Demonstration (PbD) approach for creating instructions for assistive systems. In the examination phase, we implemented the PbD approach for teaching work steps to an assistive system. After implementing the PbD approach, we conducted a first lab study for evaluating the internal validity comparing video instructions and in-situ instructions. Later, we conducted a controlled field study for learning about the effects when assembling real products in industrial manufacturing scenarios. The results of the user studies conducted in phase 2 are presented in Chapter 10.

Phase 3: Long-term effects and transfer to order picking In the third and final phase of the research conducted in this thesis, we address the question which long-term effects result of using in-situ instructions during assembly tasks over a longer period of time. To address this question, we first analyzed the requirements for long-term usage of assembly instructions through conducting interviews with experts and reviewing related work about long-term usage of instructions. As a second step, we also reviewed instruction systems for supporting workers during order picking tasks. Considering the design, we had to generalize the concepts behind our assistive system, transfer them to the domain of order picking, and add adaptivity components for enabling a long-term usage. In the examination

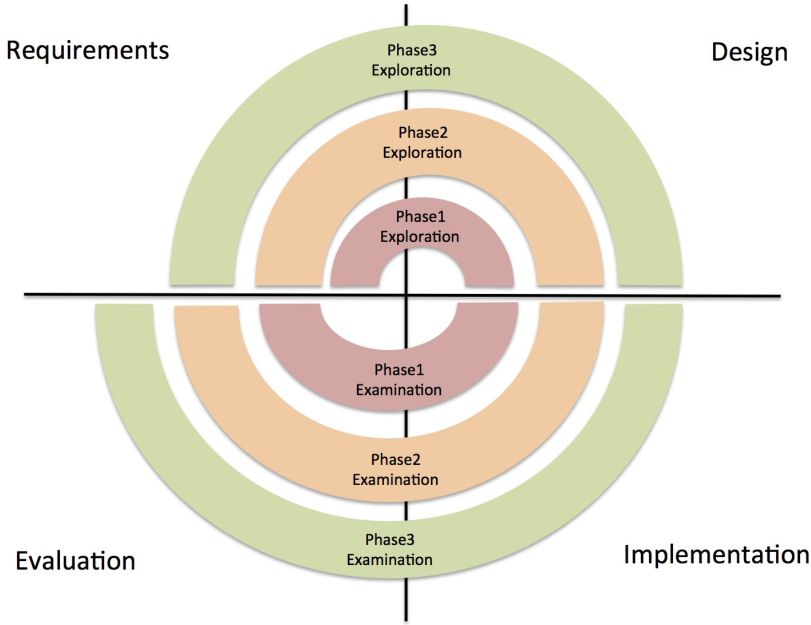


Figure 2.1: A graphical representation of the approach used for conducting research in this thesis. We use a bottom-up approach following the user-centered design process consisting of three iterations.

part of this phase, we implemented two order picking systems that follow the general concept that was developed in the design phase: a cart-mounted system and a body-worn system. Further, we implemented the concept that enables switching the adaptivity level of instructions upon correct or incorrect assembly. For evaluating both order picking systems, we conducted two laboratory studies, where we compared the systems to state-of-the-art order picking instruction systems. Further, we were conducting a long-term study in a real assembly scenario, where we evaluated the long-term effects of in-situ instructions on experienced and inexperienced workers. The results of the experiments that were conducted in phase 3 are reported in Chapter 11 and Chapter 12.

2.5 Commercial Systems

Teaching workers how to assemble products is a reoccurring task in companies with manual manufacturing workplaces. Traditionally, experienced workers are teaching the assembly steps to inexperienced workers by demonstrating them directly at the workplace. To reduce the workload of experienced workers, there are commercially available systems that can automatically teach work steps to inexperienced workers.

One of these systems are the so-called **utility videos**. These are teaching videos showing assembly steps which are recorded from a worker's point of view. Using utility videos, workers learn where to pick parts that are used for the assembly, how to hold a work piece in an ergonomically correct way, how to use and operate tools, and how to perform assembly steps. One company that is providing utility videos is for example the memex GmbH[6]. They are providing a service for producing utility videos for companies' work processes.

Another category of assistive systems for the workplace are systems that are using a **workplace-mounted monitor** to provide information according to the performed work task. These systems provide context-sensitive instructions according to the currently performed work step. To be aware of the current work step, these systems have to track the user and the performed work steps. One example of an assistive system using a workplace-mounted monitor is QualityAssist of the Sarissa GmbH[7]. QualityAssist uses a positioning system that is based on ultrasonic waves. Thereby, the user is wearing a transmitter, which emits ultrasonic waves. The waves are received and processed by a workplace-mounted system which can calculate the position of the user's hand within an accuracy of 10 mm (according to the QualityAssist website). Once a workflow was taught to the QualityAssist system, it can be played back. Through tracking the user's hand, the system can track the process of the assembly and react to forgotten work steps by displaying an error on the screen and providing an audio error signal. Other systems are using a camera-based approach for identifying work steps optically. For example the Schlauer Klaus system of the OPTIMUM datamanagement solutions GmbH[8] is using a high precission RGB-camera to optically identify and track assembly parts. They can further use the top-mounted camera to check

6 memex GmbH - `http://memex-academy.eu/` - (last access 5th Oct. 2016)

7 Sarissa GmbH - `http://www.sarissa.de/en/local-positioning-system/qualityassist.html` - (last access 5th Oct. 2016)

8 Schlauer Klaus - `http://www.optimum-gmbh.de/der-schlaue-klaus.html` - (last access 5th Oct. 2016)

if a part was correctly assembled or to verify that a part is within the tolerance. To achieve this precision, the workplace has to be equipped with a strong lighting source and requires a constant level of light at the workplace.

The third category of assistive systems are **in-situ projection** systems. These systems are using a top-mounted projector to display feedback directly on the assembly area. This approach does not require an additional screen as the feedback is directly displayed where it is needed. One of these systems is the Werklicht Pro from EXTEND3D[9]. They are using a laser-projector to highlight assembly positions, drilling positions, or providing a template for cutting. Werklicht is also able to create a documentation of performed work steps by taking a picture after the work is complete. Another system using in-situ projection is the Light Guide Systems Pro from OPS solutions[10]. In this system, a DLP projector displays picking information and assembly instructions directly at the workplace. The instructions can either be advanced manually, or can be combined with a computer vision system that only advances the instructions once the work step has been correctly performed. Other than the previously introduced assistive systems, the cubu:S system by Schnaithmann Maschinenbau GmbH[11] uses a depth camera to detect if a part was correctly picked from a box or if a part was correctly assembled at a workpiece carrier. Using a depth-camera instead of an RGB-camera makes the assembly detection and pick detection independent from the lighting conditions at the workplace. In a proof-of-concept evaluation Kölz et al. [92] used their system in a study with 20 cognitively impaired workers and found that their system increased the participants' motivation.

[9] EXTEND3D - `http://www.extend3d.de/werklichtpro.php` - (last access 5th Oct. 2016)

[10] OPS solutions - `http://www.ops-solutions.com/products.html` - (last access 5th Oct. 2016)

[11] cubu:S - `http://www.schnaithmann.de/fileadmin/user_upload/PDF/Schnaithmann_ASSISTENZS_D.pdf` - (last access 5th Oct. 2016)

Chapter 3

Related Work

Related work that is relevant for assisting workers during manual assembly work tasks is based in different areas of research in human-computer interaction, psychology, and mechanical engineering. In this chapter, relevant related approaches are described and are assigned to the following categories: Augmented Reality, sensing interaction, assistive systems at workplaces, instruction giving in assembly, and order picking.

3.1 Augmented Reality

Augmenting reality with information goes back to Sutherland [147]. In his prototype, he overlayed the view of participants with objects that are close and objects that appear to be far away. According to the reality-virtuality continuum by Milgram and Kishino [115], Augmented Reality is starting from real environments and successively adding additional information into a real world scene. This overlaying of real world scenes can be applied to any scene in any context. E.g. in a desktop scenario, Wellner's DigitalDesk was the first system that combined a camera and a projector for creating an augmented table that can merge digital information with physical objects that are placed on a desk. The DigitalDesk uses an RGB camera to detect the position of a paper on the desk and to detect where a user is pointing. Further, a projector is used to highlight information directly on papers that are placed on the desk.

However, the idea of augmenting work processes with visual information has been around about only two decades. In 1993, Caudell and Mizell [32] suggested using HMDs for displaying drilling spots and instructions for a manufacturing task inspired by aircraft manufacturing. Over the years, research has defined sub-categories of Augmented Reality according to the different use cases and the ways of presenting information. For example, we refer to SAR [130] when an object is being displayed directly onto the physical space around the user. An example for SAR is the Everywhere Displays Projector [126], where information is projected directly into the physical world with respect to the physical properties. In this project, a projector and a rotatable mirror were used to augment physical objects with digital information. As distortion was a problem, Pinhanez suggested to correct the projection using a camera to enable a distortion free projection on curved surfaces.

Another sub-category of AR is IAR, which refers to using Augmented Reality for industrial use cases. Navab [117] and Fite-Georgel [43] categorize IAR according to the different use cases: product design, commissioning, training, manufacturing, inspection and maintenance, and decommissioning. Comprehensive surveys [118, 132] show that IAR can be used to support almost every aspect of a manufactured product's life-cycle. Experiencing a designed product can be done immediately [42], industrial robots cooperating with human workers can be programmed using AR-debugging approaches [33], order picking can be supported using HMDs [70], and maintaining existing machines and products can be supported directly on site [160]. Workers can even be motivated during the work tasks by using IAR for gamification [93]. For production planning purposes, Otto et al. [122] proposed using floor projection to directly overlay the shop floor with information from a digital planning system. Furthermore, in the domain of quality control, Nolle and Klinker [119] proposed using CAD data of manufactured products and overlaying the reality with them. Zhou et al. [166] use in-situ projection for highlighting welding spots in manual welding tasks for quality control after and during a welding task. They experiment with different visualizations of the highlighted spot using a stationary projector.

While the most previous projects mainly are designed for stationary setups, Beardsley et al. [18] use a mobile camera-projector system. They are using their system to align the projection next to unique points in the environment. Thereby, the projection can stay at a defined position even when moving the projector. Further, they were able to use the movement of the hand-held projector as a digital cursor for interaction. Willis et al. [156] scaled up using a projector as an input device to using multiple mobile projections that are interacting with each other. They use an IR channel to communicate between multiple mobile camera-projector pairs. Moreover, Raskar et al. [129] created a geometrically aware camera-projector system by adding a tilt sensor and creating a 3D-mesh from the camera input. Projected images are then transformed according to the 3D-mesh and corrected to be viewed distortion free even on non-planar surfaces. They further used their system to project feedback onto picking bins, however their focus was mainly on distortion free viewing and combining multiple projectors to enable a projection in 3D space. Schwerdtfeger et al. [142] are using head-mounted and environment-mounted laser projectors to display information in a welding context. Their findings comprise that head-mounted projectors are too heavy to use in long-term tasks e.g. at workplaces. Löchtefeld et al. [109] use a hand-held camera-projector system that can be directed at a shelf to categorize products according to a user's personal profile. They are using an RGB-camera to visually identify objects in the shelf and augment them with information. Harrison et al. [76] targeted a mobile scenario by mounting a camera-projector pair on a user's shoulder. Recently, Winkler et al. [159] proposed a backpack-mounted solution for both projector and depth-sensing camera in their AMP-D project. Especially for interacting with mobile projectors, many areas of application have been suggested. Rukzio et al. [135] and Wolf et al. [161] provide a comprehensive overview. An overview about placing a body-worn projector is provided by Ota et al. [121].

Also in other domains, in-situ projection has already been used to teach and instruct learners. For example in the domain of learning how to play instruments, Weing et al. [153] are using in-situ projection on a piano to support learners in playing the piano. Similarly, Löchtefeld et al. [108] uses mobile projection to support guitar learning.

3.2 Sensing Interaction

Apart from presenting information, sensing interaction is the most important aspect of building interactive systems. Traditionally, interactive systems are operated by using graphical user interfaces (GUIs) following the WIMP (window, icon, menu, pointing device) paradigm. However, more recently user interfaces for interactive systems have been proposed to be made tangible [85] or additionally allow for being operated using natural interaction [86].

A major part of creating a natural user interface is the possibility to recognize gestures. For detecting 2D gestures, touch events need to be detected by a surface. An example is the Touchlight system [157], which uses two RGB-cameras to detect touch input on a projected surface. Thereby, a user is able to interact with projected content. With the proliferation of Kinect_v1 depth cameras in 2010, sensing touch on projected interfaces became easily possible on arbitrary surfaces. Therefore, Wilson [158] suggested an algorithm that observes the depth data in close proximity to surfaces. Whenever, a user touches a surface with a finger, a touch event can be detected. Later, Wilson's algorithm was improved by Hardy et al. [75] by using KD-trees to handle multitouch with 30 frames per seconds. Further, the dSensingNI project [90] combines Wilson's algorithm with gestures in a tabletop system. They also support detecting the presence, volume, and orientation of cubical objects using a top-mounted Kinect_v1 depth camera. As their algorithm uses multiple depth layers, it is even capable of detecting touch on physical objects that were placed on the tabletop and detecting stacked objects.

For detecting gestures in 3D space, a user usually has to be equipped with a sensor or carry a sensor. E.g. Schlömer et al. [140] are detecting 3D gestures using a user-carried sensor. However, in the domain of assembly detecting 3D gestures with their full trajectory is not needed for interacting with assistive systems. For example Bannat et al. [13] present a framework using a top-mounted RGB camera to detect bins automatically based on their color and shape. Once the position of the bins is known, their system uses the RGB-camera to detect the position of the worker's hand. Thereby, the 3D movement of the worker's hands is simplified to just using the current position of the hand and defining interactive zones. In their system, assembly instructions are shown on a monitor close to the work area. The system highlights the next bins to pick from using a top-mounted projector. Korn et al. [98] extend this approach by using a top-mounted depth camera instead of an RGB camera and a top-mounted projector in production environments. The position of the bins and the position of an assembled part have to be defined manually using a graphical editor. Their system then highlights the bin to pick from. As their system cannot automatically detect the correct assembly in each step, it uses projected buttons that the user can manually advance the projection to next steps. This technology was extended by Rädle et al. [128], who equipped a lamp with a depth camera. With their algorithm they can track hands and objects and thereby enable multi-user interaction on non-instrumented tabletops.

Another strand of work has focused on augmenting parts of products or tools with sensors and sense interaction using the sensor data. For example, Antifakos et al. [7] use instrumented tools and instrumented assembly parts to infer a user's current action. Based on the action they proactively suggest instructions for assembling an IKEA PAX wardrobe. Compared to a printed manual, their system can dynamically react upon users' actions as it is aware of all possible assembly orders rather than having one programmed and fixed order. Alm et al. [6] suggested using a smart assembly trolley that is equipped with force sensors and RFID readers to infer the actions of the worker. Based on the sensor data the assembly trolley is able to check if a part has been taken out, placed on the trolley, or if a tool is being used. Moreover, Knibbe et al. [91] are equipping tools with sensors and connect them to a smart makerspace, which provides multimodal instructions for assembling DIY projects based on the state of the tools and the workflow of the project.

Instead of augmenting the assembly parts, other research proposed mobile systems for displaying interactive assembly instructions by augmenting the users with sensors. For example, Ward et al. [151] equip the user with body worn microphones and accelerometers to infer the user's current activity in an assembly environment. Even when combining multiple features [24] to recognize an activity more reliably, a body-worn system unfortunately cannot detect if a part is assembled correctly. More recently, Kritzler et al. [102] use RFID technology and a smart watch to ensure that a worker is wearing safety equipment before starting a task. By equipping each piece of safety equipment with an RFID beacon, a signal from all beacons is only received by the smart-watch when the worker is wearing all pieces of the safety equipment. Using this concept, a worker can only enable a machine when every piece of safety equipment is worn.

Sensed interaction cannot only be used to interact with an existing system, it also can be used to teach or program new workflows or procedures. Often this act of showing an interactive system how a task is performed is called Programming by Demonstration (PbD) (also referred to as programming by example). PbD was initially proposed to enable users to record macros without knowing any programming language or writing code. This approach has been adopted by many application domains which comprise desktop applications like MS Excel, computer-aided design, and text editing [106]. Thereby, a user's actions are translated into a textual procedure, which later can be played back and altered. For example, the Peridot system [116] enables interface designers to demonstrate how a UI should look like rather than having to program it. Recently, Kubitza and Schmidt [103] introduced a framework that enables non-programmers to use PbD to program for smart environments.

Further, the PbD approach is also used to teach new motion sequences to humanoid robots by recording movements of a human worker. Aleotti et al. [5] reproduce and optimize measured trajectories of a human worker. The trajectories can then be used to infer high-level actions [21]. After defining actions, the sequence of the actions can be played back and altered. Instead of programming physical robots, Marinos et al. [111] use a PbD approach to rapidly create animations for a virtual robot inside a blue or green box of a virtual studio.

Overall, previous work on sensing interaction uses either 2D surfaces for detecting gestures that are performed on the surface, uses body-worn or carried sensors to detect gestures that are performed in 3D space, or creates simplified abstractions from 3D trajectories. The sensed interaction is not only used to directly interact with systems but can also be used to sense generic actions and use them for teaching workflows to a system.

3.3 Assistive Systems at the Workplace

Assistive systems for workplaces have been proposed to facilitate collaborative work, provide a continuous support for persons with disabilities, and for providing cognitive assistance during complex tasks.

To facilitate the helping process for work tasks among peers, McCalla et al. [114] created an assistive system. Their system provides a database containing information about which colleague knows which processes. If colleagues stated that they are willing to help others, the system does an automatic matching of colleagues that are willing to help and the tasks that a user needs help with. Another approach for an assistive system facilitating collaborative work is the TeleAdvisor system [72], which uses a camera projector system to enable a remote helper to give instructions to an on-site learner. They use a task where workers are connecting cables at a TV setup scenario. Furthermore, Gauglitz et al. [66] use a hand-held device that can be annotated with AR-instructions for remote collaboration in a Boeing 737 cockpit scenario. In contrast, Sakata et al. [138] use Lego Duplo bricks for giving instructions for assembly tasks in a remote collaboration scenario. Another approach is presented by the T.A.C. system of Bottecchia et al. [26], which supports a remote expert who can augment the view of a worker with 3D AR-elements which are displayed in an HMD.

Considering assistive technology for persons with cognitive disabilities, a general literature review is provided by Sauer et al. [139]. They conclude that especially for this target group of persons with cognitive disabilities, assistive technology can have a positive effect on the person's performance and therefore enable building more complex products.

Assistive technology at the workplace for providing cognitive assistance during complex tasks are implemented using many different technologies. One of these technologies is presenting assembly instructions on a mobile display. These mobile displays are either carried or worn by workers during assembly tasks. Echtler et al. [39] use a display that is mounted directly at a welding gun to provide information about the position of welding spots. In their work, they are using a tracking system that is mounted in the work environment to track the position of the welding gun at all times. Thereby, it is possible to show the exact welding position to the worker on the display. Other work suggested presenting assembly instructions using a chest-worn display (CWD) [138], a nearby screen [71, 99], using a mobile phone [22], or a tablet computer [67]. Gavish et al. [67] compare tablet-based AR instructions to interactive VR instructions. They found that AR and VR training requires a longer TCT than video-based training. A CWD is used by Sakata et al. [138] as they compared it to an HMD in a remote collaboration assembly scenario. In a user study, they found that the CWD is more suitable for the task compared to the HMD. Moreover, Aehnelt and Urban [2] suggest using a combination of stationary displays, mobile displays, and private displays. Especially the private displays can be designed in a user-worn way using a smartwatch to provide instructions to workers.

Considering stationary displays, Korn et al. [99] conducted a study with 81 impaired workers, where they compared in-situ pictorial instructions to instructions that are presented on a nearby screen. They found that pictorial in-situ instructions lead to a faster assembly, but workers were making more errors compared to the on-screen control condition. In their study, the participants assembled on average 23.6 minutes using the pictorial in-situ instructions. Also Marner et al. [112] compared in-situ projected instructions to instructions that are shown on a screen. They conclude that in-situ instructions are faster and lead to less errors.

Other assistive systems using AR for manufacturing are presenting instructions on HMDs. For example Tang et al. [149] showed that spatially overlaying the assembly workplace with AR instructions using an HMD reduces the error rate in assembly tasks by 82% compared to paper-based instructions, instructions on a monitor, or instructions that are steadily displayed on an HMD. In their study, they introduced an abstract pick-and-place task using Lego Duplo bricks. Further, Hahn et al. [73] are using see-through HMDs for highlighting picking positions and assembly positions and displaying textual instructions in assembling a printed circuit board. In their study, they are measuring how confident workers are when using their HMD-based approach. Similarly, Paelke [123] uses a spatially registered AR feedback that is displayed on an HMD for providing assembly instructions in a smart assembly workplace. Moreover, through a user study Henderson et al. [80] report that users have less head movements using HMD-based AR instructions while repairing a vehicle. More recently, Zheng et al. [165] provided further research towards finding the optimal position for displaying feedback on an HMD. For evaluating the position, they conducted a car maintenance task – checking a car's oil level and changing a light bulb. In their study, they compare providing instructions using a central position on the HMD, which is directly in the user's field of view, against a peripheral position, a hand-held tablet representation, and printed paper instructions. Their results reveal that a central HMD representation is faster than the peripheral representation. Further, they did not find a difference in completion time between the HMD and non-HMD approaches.

Henderson and Feiner [81] use an HMD to display three-dimensional arrows and text-instructions for maintenance tasks of an armored personnel carrier turret. In their user study, they found that users could locate the tasks more quickly when using the HMD. To interact with the instructions the user operates a wrist-worn controller that advances the feedback. The qualitative feedback indicates that the expert users liked the AR-approach. They were using the HMD in the study for approximately 75 minutes. Furthermore, De Crescenzio et al. [36] use 3D elements for maintenance tasks and introduce a marker-less tracking of the HMD. This approach was further investigated by Henderson et al. [82], as they compared 3D elements that are presented on an HMD to instructions that are presented on a monitor. They found that using 3D elements on an HMD leads to a faster and more accurate assembly. Yuan et al. [163] are proposing to use a combination of an HMD and a pen for presenting instructions during assembly tasks. Instructions are presented on the HMD as an annotated picture showing the correct assembly process and textual information. Their system does not require the assembly components to be tracked, as the HMD tracks the pen that is carried by the user. Once the user wants to advance the assembly instructions or interact with the system, the pen can be used as a cursor.

HMDs are also used in other areas than an assembly workplace, e.g. Zauner et al. [164] use AR markers to provide assembly instructions on an HMD for assembling furniture. Further, Kim and Jun [89] use an HMD and a camera to track a user's location based on the images from the camera and then displaying navigation information on an HMD.

Other assistive systems focus on using in-situ projection to display assistance directly onto the workplace. Early versions of in-situ projection systems were presented in 2003 by Sakata et al. [137]. They introduce a wearable active camera/laser-pointer that a remote expert can control to provide assembly help. With increasing technology, an assistive system using a top-mounted projector and a top-mounted camera was introduced by Bannat et al. [14]. They use an

RGB-camera to detect which bin the worker is picking the next part from. Their system can provide context-sensitive help at the workplace according to which assembly part was picked by equipping the worker with a grasping sensor. This grasping sensor ensures that the worker actually picked up an item from the bin and that the sensor did not just register the placement of the worker's hand above the bin. Rüther et al. [136] use projection for displaying information in sterile environments. In their study they found that using projected instructions for cleaning medical instruments is well received and leads to less errors than using paper-based instructions. In 2012, Korn et al. [97] suggested using motion and voice input for sensing and triggering events at an augmented workplace using in-situ projection. They further suggested using gamification elements in conjunction with the measured interaction to motivate workers during their work tasks. Recently, Büttner et al. [30] presented an assistive system using a top-mounted projector which is displaying picking information directly onto the bins where the parts have to be picked from. As in their use case, the workers are performing all assembly steps in their hands without using a workpiece carrier, the assembly instructions are projected onto an instruction area at the workplace. Further, their system provides a foot pedal, which the worker can press to advance to the next work step.

3.4 Instruction Giving in Assembly

When designing assistive systems for providing instructions the design of presented instructions is very important. Projects focused on how to generally visualize instructions, focused on how to present instructions for persons with impairments, and compare different ways of presenting instructions to each other. E.g. in 1999, Boud et al. [27] compared AR and VR instructions to traditional 2D drawings and found that the interactive instructions out-perform the traditional ones.

One of the ways to provide instructions for manual assembly tasks is using textual descriptions. In 1992, Cuvo et al. [35] used textual instructions to teach tasks to persons with mild cognitive disabilities. They found, that feedback about the performance is important for the workers. More recently, the LuminAR [107] system included a camera and a projector in an anglepoise lamp. When objects are placed under the lamp, they can be augmented with additional textual information. This technique can be used in a shop window, showing the price and information about exhibited articles. In the LuminAR system, text and images are used to display this information.

Compared to text, pictorial instructions are more widespread as they are language independent and do not require the user to be able to read. Pictorial instructions are used for teaching daily life skills to persons with cognitive disabilities e.g. how to dress themselves [125], how to clean, or how to cook [87]. In a study, Steed et al. [146] investigated if persons with cognitive disabilities are capable of learning daily life tasks without being continuously supervised by a socio-educational instructor. After an initial instruction how to use pictorial instructions, the participants had to learn how to use a vacuum cleaner just using the pictorial instructions. The results show that using pictorial instructions, the participants were significantly better. Additionally, participants could remember the instructions over a longer period of time and even learn new tasks by just using pictorial instructions. Lancioni et al. [104] experimented with pictorial instructions for performing tasks. In a study, they compared instructions on an computer-aided palm device to instructions on cards. Participants using the computer-aided palm device to view the pictorial instructions performed better. Also considering the subjective feedback of the participants, the computer-aided instructions were preferred. Korn et al. [98] and Bannat et al. [14] also use a camera-projector system that uses a similar concept as our assistive system. In their systems, both use pictorial instructions in a manufacturing environment for assembling LEGO models. The images used in their projected instructions look exactly as the ones in printed manuals. Moreover, Hashimoto and Siio [78] are using in-situ projected contour visualizations of Lego bricks as assembly visualization. Another strand of research focused on how to build easily understandable pictorial instructions [3, 79]. There, research suggested building hierarchical pictorial instructions where the reader can see the action that is being performed. Step-by-step instructions enable the reader to better identify the step that is being performed. Furthermore, the parts should be oriented in a way that all important features are visible to the reader. A number of pictorial instructions are built to comply with the design guidelines proposed by Agrawala et al. [3].

Considering video-based instructions, Rüther et al. [136] use video-based in interactive in-situ instructions using a projected user interface. Their system uses a camera-projector system to provide interactive instructions in a sterile area. Moreover, Suzuki et al. [148] use in-situ projection for displaying the hand movements of expert workers. Thereby novice workers can learn assembly tasks by mimicking the hand movements of expert workers.

When looking at assistive systems that have been introduced, the feedback that is provided by the assistive systems is mostly visual. However, there are systems that are providing additional auditory and haptic feedback. Auditory feedback during assembly tasks is proposed by Rauterberg and Styger [131]. In their study they used a simulated maintenance task for assembly lines to compare visual feedback against visual and auditory feedback. Their results suggest that additional auditory feedback improves the performance of the users and leads to a more positive mood. Haptic feedback has mostly been suggested for situations where visual or auditory feedback is not appropriate or dangerous. E.g. Bial et al. [20] proposed using a glove for providing vibrotactile feedback for navigating motorcyclists while driving. Haptic, visual, and auditory feedback have been compared for a variety of tasks. E.g. Akamatsu et al. [4] compared the three modalities as a feedback for pointing tasks. Although no differences in task performance were found, the authors discovered important design implications considering the choice of the feedback modality. Moreover, Richard et al. [134] investigated the effect of haptic, auditory, and visual feedback on manipulating virtual objects. Their results suggest that both auditory and haptic feedback improves the operators' performance. The three feedback modalities were also compared in a telepresence assembly task by Petzold et al. [124]. In their study, they found that additional haptic feedback significantly increases the work effectiveness of operations.

3.5 Order Picking

Systems for supporting workers during order picking tasks and systems supporting users in finding objects have been the topic of various projects. A strand of work has embedded cameras and projectors to the environment to track the location of objects and provide visual feedback. Butz et al. [31] use a stationary camera-projector system, which is firmly mounted at a room's ceiling. Their system can automatically detect books that are equipped with visual markers and later highlight their position using a projector. Furthermore, Crasto et al. [34] use a foreground detection algorithm to sense changes in a bookshelf. Li et al. [105] are using a stationary Kinect_v1 together with computer vision algorithms to identify picked objects based on their shape and visual appearance. In their approach, the worker has to explicitly place the object in front of the camera which results in an extra work step and might increase the TCT. Recently, Bächler et al. [11] investigated how beneficial an order picking system using in-situ projection might be for workers with cognitive impairments. Their results reveal that 85.9% of the interviewed persons benefit from an interactive system. Further, they evaluated different pictogram visualizations for order picking tasks, which are suitable for cognitively impaired workers [10]. In a comparative wizard of oz study, Bächler et al. [9] evaluated four picking visualizations: Pick-by-Projection, Pick-by-Light, Pick-by-Display and a Pick-by-Paper baseline. Their results reveal that the Pick-by-Light method was significantly faster than the other methods used in their study.

Other projects are using HMDs to deliver feedback directly to users. For example, Reif et al. [133] and Schwerdtfeger et al. [143] use a modification of the attention funnel, proposed by Biocca et al. [23], on HMDs to guide workers to the next shelving unit in a warehouse. They use an optical tracking system to display the funnel depending on the user's position and orientation on an HMD. It is reported that the ER can be reduced using the attention funnel visualization. However, the authors also report that the weight of the HMD can distract the worker, the visual clutter produced by the funnel limits the field of view, and the precise position of the HMD is crucial for the visualization. Furthermore, content that is displayed on an HMD could potentially block the safety-critical real-world view [84]. This makes it difficult to deploy in industrial settings. To overcome the need for a precise location of the HMD, Weaver et al. [152] suggest to display a 2D model of the shelf on the HMD and highlight the bin to pick from in the model. Guo et al. [70] extend this approach with an intelligent batch picking technique that allows for picking multiple orders at the same time. They compare this approach to a cart-mounted display (CMD), which is displaying the 2D graphical representation and the order picking tasks. In their results, the 2D representation that is displayed in the HMD is significantly faster than the traditional Pick-by-Paper (PbP). Furthermore, their analysis also shows that schematic representation of the warehouse on the CMD is also favored compared to a PbP approach regarding ER, TCT, and cognitive load.

The scenario of order picking is so far the only scenario where long-term evaluations of HMDs have been conducted. For example, Schwerdtfeger et al. [143] tested their approach in a two hours study to get insights about long-term usage of AR in production environments. After using the HMD for two hours participants reported headaches, problems to focus on the instructions shown on the HMD, and needed a 15-minute break from using the HMD. Grubert et al. [69] report another long-term evaluation of order picking processes with a four-hour duration. They could reproduce Schwerdtfeger et al.'s findings. Moreover, Tumler et al. [150] analyzed the physical effects of long-term usage of an HMD in an order picking scenario. However, they could not find a difference in the users' strain between HMD and traditional paper-picking.

III

SYSTEM

Chapter 4

Requirements and Concept

This chapter is partly based on the following publications:

- M. Funk, O. Korn, and A. Schmidt. An augmented workplace for enabling user-defined tangibles. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, pages 1285–1290. ACM, 2014
- M. Funk and A. Schmidt. Cognitive assistance in the workplace. *Pervasive Computing, IEEE*, 14(3):53–55, July 2015

We are designing and implementing an assistive system according to a user-centered design process. Therefore, we first identify requirements and define interaction concepts. The requirements and concepts that are introduced in this chapter are technology independent and do not rely on the underlying hardware of the assistive system.

To support workers adequately at the workplace with an assistive system, the system has to know details about the worker's experience, the performed task, and the best ways to provide help for the performed task. For example, a non-experienced worker might need more support than a more experienced worker during the task. Therefore, the system needs to decide if a worker is more experienced or if the worker still needs support. This could be achieved by detecting if the worker is making errors or not. Thus, the system needs to be able to distinguish between correctly performed work steps and being able to detect errors. The fact that an assistive system can detect correctly and incorrectly performed steps might be one of the biggest benefits of an assistive system. It takes the decision whether a task was correctly performed or not away from the worker and decides automatically about the completion of the task. This should reduce the cognitive effort that is needed by the workers at the workplace.

The use of an assistive system at the workplace should not result in any additional effort. For example, using the system should not result in any additional actions and a user should not need any specific instruction for handling it. Moreover, the worker should be able to interact with the assistive system implicitly, just by performing the task. This means that the system needs to be aware of the context and the actions that are being performed at the workplace by using different sensor data (cf. [141]).

Considering the presentation of the instructions, the assistive system should provide instructions that are easy to understand. Further, understanding the instructions should result in the least possible cognitive effort.

The assistive system should implement a **tri-state assembly logic** for presenting feedback, consisting of three states: *correct*, *default*, and *error*. In the *correct* state, the work step was performed correctly. Usually a correctly performed work step results in advancing the feedback to the next work step. The system is in the *error* state if a mistake was detected (e.g. picking a wrong part or assembling a part at the wrong place). Usually, in the *error* state, the system is showing error feedback to help the worker to correct the error. If the system is neither in the *error* state nor in the *correct* state, it should be in the *default* state. In this state, the assembly is not yet finished and no error is detected. This could mean that the user is currently working on the task or is not working on the task and is making a break. In the *default* state, the instruction how to perform the task should be presented.

4.1 Use Cases

Assistive systems might have a great potential for different areas of applications in the whole process of manual assembly. Therefore, we define use cases (U), which we will investigate in this thesis and which would benefit from using an assistive system.

U1: Training workers for a new task
Assistive systems can be used to train new workers or already experienced workers in a learning how to conduct a new task. Instead of learning from another worker, the assistive system teaches the new worker directly on the task by providing context-aware instructions and detecting and reacting upon errors that were made.

U2: Continuous quality support
Another use case for assistive systems at the workplace is a continuous support of workers during a task. This could be beneficial if the task is very cognitively demanding (e.g. when producing in lot size one) or when a worker needs a continuous support (e.g. when supporting cognitively impaired workers).

4.2 Requirements

We define requirements (R) for an assistive system for the workplace. These requirements are either functional requirements, which describe a function that is necessary for the assistive system to work, or non-functional requirements that cannot be measured directly but are very important for the maintenance and long-term usage of the assistive system. In 2014, Korn [93] already outlined requirements for context-aware assistive systems to enable gamification and quantified self at the workplace. In this work, we extend Korn's requirements and define additional requirements from a systems point of view.

4.2.1 Functional Requirements

R1: Detecting picked parts
Picking parts is an activity that needs to be performed at many workplaces where manual assembly tasks are done. Parts are usually stored in picking bins, shelves, or compartments. An assistive system would need to detect when a worker is picking a part.

R2: Detecting incorrectly picked parts
Similar to **R1**, an assistive system should also detect if a worker is picking a wrong part from the storage place. The system should warn the worker when it detects that a wrong part is picked. Thereby, an error can be prevented before a wrong part is assembled. This is especially useful if a task consists of many parts that are similar and can be easily mixed up.

R3: Detecting correct assembly
According to the **tri-state assembly logic**, the system needs to detect when a task was correctly performed. One of the tasks at an assembly workplace is the assembly of parts. Thus, a key requirement for the assistive system is to be able to detect the correct assembly of parts. Thereby, the system needs to check if the assembly is in its defined final position after each work step. Only when correct assembly can be detected, the system can enter the *correct* state and advance the provided feedback after a step has been performed correctly.

R4: Detecting assembly errors
According to previously presented **tri-state assembly logic**, a work step is not advanced until the system can detect that it was performed correctly. As the same is true for presenting error feedback, the system has to be able to detect assembly errors. Thereby the system can present error-feedback when a mistake was made.

R5: Detecting usage of tools
In many work tasks the usage of tools is an important step of the workflow. As using the right tool on the right part of the assembly is very important and a workplace could feature different tools that can be applied to the same part, the system should also guide the worker in using tools. Therefore, the system should at least be able to detect if a user is picking the correct tool from its defined place.

R6: Provide understandable instructions
The goal of an assistive system is to provide instructions at the workplace to help workers performing the tasks. The presented instructions should be easily understandable, targeting the currently performed task, should be context-sensitive i.e. reacting upon a workers actions, and resulting in very little additional cognitive effort.

R7: Communicating with other devices
An assistive system is running according to its defined workflow. However, in many cases other components have to be synchronized with the workflow of the assistive system. Therefore, the system should implement a possibility to communicate the current state and information about actions that were performed at the workplace, feedback, errors, and further states to other devices. Accordingly the external devices can provide additional information that are used by the system (e.g. additional sensor data or external commands).

R8: Enable different levels of details in instructions
As there is a difference in the experience levels of the users, an assistive system needs to provide different levels of details in the instructions. For instance in **U1**, the users are not experienced with the task as they are learning the task

using the assistive system. A more detailed explanation would be required in this use case. In contrast, workers that are using the system in **U2**, might not need feedback containing all details. They might just need a hint, which is telling them which step to perform next, or which part to assemble next, as they already know tasks that need to be performed. This is especially useful when producing many variants of a part and the order of producing the parts is not always the same.

4.2.2 Non-Functional Requirements

In addition to the previously stated functional requirements the system should include four non-functional requirements. First, the system should run stable and reliably as it will be deployed in an assembly environment. Depending on the shift plan of the factory where the assistive system is used, the minimum run time of an assistive system will be eight hours per workday, i.e. one shift per day. As some factories produce in up to three shifts per day, the assistive system needs to be available up to 24 hours per day. Another requirement is that the system should not require being re-calibrated during the use. Once a calibration was stored, it should be valid until a parameter was changed. Considering maintenance, the assistive system should be easy to maintain. This requires an intuitive user interface and defined points where extensions can be integrated.

4.3 General Architecture

According to the previously outlined requirements for assistive systems, we introduce a general architecture for implementing an assistive system. Figure 4.1 shows an overview about the components. We distinguish between three types of components: activity recognition components, feedback components, and logical components. The activity recognition components encompass a pick detection (**R1, R2**), an assembly detection (**R3, R4**), and a tool detection (**R5**). The feedback components include providing understandable feedback (**R6**) and a mechanism to enable adaptive feedback that is tailored to the users' skills and performance (**R8**). Lastly, the logical components include defining a workflow and enable communication with other devices (**R7**). This architecture combines all these components creating the building blocks for an assistive system that fulfills the requirements. As for now the building blocks are generic, independent from the use case, and independent from the underlying technology.

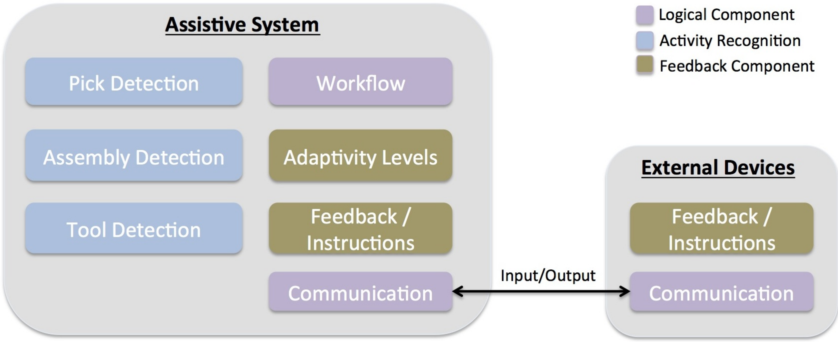


Figure 4.1: A general architecture including building blocks that fulfill the outlined requirements for an assistive system. The architecture consists of activity recognition components, logical components, and feedback components.

4.4 Interaction Concepts

The previously introduced requirements and the general architecture allow for using the activity recognition components as an input for interacting with the assistive system. Overall, the design goal of the system is to use as many natural interaction as possible and only require a GUI for configuration. In the following, we describe three interaction concepts that we implement in our assistive system: Programming by Demonstration, implicit interaction with instructions, and the use of user-defined tangibles.

4.4.1 Programming by Demonstration

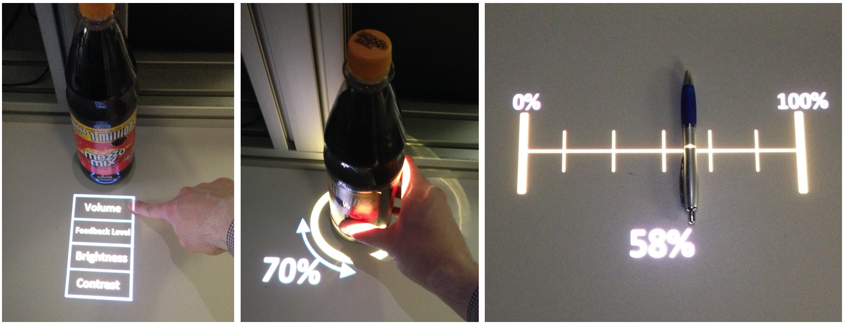
One important aspect of assistive systems which are providing instructions for an assembly workplace is how instructions can be created. Inspired by related work in Programming by Demonstration, we are applying the concept of demonstrating a workflow to an assistive system – just by performing it once. The system should be able to create instructions for the performed work steps according to the performed actions at the workplace. Thereby, we are using the activity recognition components that were outlined in the architecture previously (cf. Figure 4.1): pick detection, assembly detection, and tool detection. As for using the assembly detection, the system has to trigger explicitly the recognition after the assembly was performed, we need to implement a feature that is able to sense when an action is still ongoing. Therefore, we implemented an additional activity recognition algorithm using the RGB image provided by the Kinect_v2. Using this algorithm, the system is able to see when a worker is standing in front of the assistive system. Our algorithm assumes that when a worker is standing in front of the system, the assembly step is still in progress. Once the assembly step was performed, our algorithm requires the worker to step out of the assembly area as the depth data that is created by the presence of the worker influences the automatic assembly detection. After the system performed the assembly detection, it gives the worker a visual signal that it is ready to record the next work step. Once all work steps of a workflow are performed, the system automatically creates an interactive assembly instruction using visual feedback.

The feedback that is automatically created by the system is different for each type of performed action. For picking a part from a bin, the automatically created instruction highlights the bin to pick from using a green light. When an assembly step was performed, the contour of the assembled part is highlighted at the position where the part was assembled. For using tools, the position of the tool is highlighted using a yellow light until the tool is put back to its distinct place.

In the end, the system creates a workflow that is in the same order as the worker was demonstrating the work steps. Afterwards, the worker can edit the automatically created workflow and enhance it with e.g. videos or textual information.

4.4.2 Implicit Interaction with Instructions

During the daily use of the assistive system, workers should interact with the system implicitly – just by performing the work steps that they would normally do without using an assistive system. This interaction concept does not require the worker to operate a complex GUI anymore as the interaction with the system is solely based on performing physical actions at the workplace. The implicit interaction concept also uses the activity recognition components that were outlined in the system's general architecture (cf. Figure 4.1). Thereby, the system can advance instructions when a worker picks from a correct bin, assembles a part correctly, or uses a correct tool. In contrast, if an error is made in the assembly or if the worker picks a part from an incorrect bin, the system is able to implicitly display an error state. The error state is exited again if the worker either resolved the assembly error or picks a part from the correct box. Using this concept, the workers are not required to acquire knowledge about operating the assistive system, as they interact only by performing actions on the workplace.



(a) Selecting a function (b) A rotary knob control (c) A slider control

Figure 4.2: (a) After registering the user-defined tangible, the user can choose a function that is assigned to the tangible control (b) Round everyday objects can be used as a rotary knob by rotating it (c) The value of the slider is selected by positioning the tangible control on the slider.

4.4.3 User-defined Tangibles

For further interaction with digital information and accessing functions of the assistive system, we propose and implement the interaction concept of *user-defined tangibles* [51]. As the assistive system enables a precise tracking of objects that are placed at the workplace, we propose to use regular physical objects that are spontaneously assigned to a digital function for triggering or controlling the assigned function. The user-defined tangibles interaction concept uses the object recognition that is implemented in the tool detection. Additionally, the concept requires a precise finger tracking. The interaction concept consists of four steps: defining tangibles, choosing and linking functions, using the tangibles, and finally unlinking the tangibles and the digital functions.

Defining Tangibles
First, an every-day object has to be made familiar with the assistive system. Therefore, the assistive system provides a learning-mode. In this mode an object is placed into a pre-defined area on the workspace. The top-mounted camera takes a picture of the object. In a second step, the depth data of the object is captured to learn the object's shape. By using the position in the captured image as a reference, the system can now track the object and determine its current position and orientation.

Choosing Functions
Once an object is registered with the system, the user can choose a function that is assigned to what has now become a user-defined tangible control. The functions are shown in a list next to the tangible object (see Figure 4.2(a)). The user can select a function by touching it.

Using the Tangible
In a next step, the user can choose the tangible control's type. The system supports using the control as a knob, which reacts to rotating (see Figure 4.2(b)), or using the control as a slider on a projected scale (see Figure 4.2(c)), which allows placing the tangible control on the desired value. As a projected scale consumes space on the table, the scale is only projected when the assigned tangible control is moved. As a workplace can be filled with many objects, our system highlights objects that are bound to a digital function by displaying a yellow circle around them. When an object is active, the yellow circle is replaced with feedback information according to the type of control (e.g. the scale in Figure 4.2(c)). If a user-defined tangible control is idle for a while, the feedback information is replaced with a yellow circle again.

When a user-defined tangible control is removed from the workplace, the control is still bound to the function. This can be used to remove currently unused controls from the work area and use them again at a later time. In a future scenario, tangible controls could also be shared across interconnected workplaces. A co-worker could ask for the brightness-tangible and could be handed a screwdriver, which was bound to controlling the brightness at a workplace nearby.

Unlinking
In order to turn a user-defined tangible into a normal object again, it has to be unlinked from the assigned function. To be robust against false-positive unlinking actions, we chose to use gestural input for unlinking objects from their assigned functions. The user can select a bound object by pointing at it using a one finger pointing gesture with both hands. This gesture was chosen because it is robust against false-positive triggering. The system then highlights the object that is marked for unlinking with a red circle. If the object is now removed from the table, the assigned function is unlinked from the object.

Chapter 5

System Implementation

In this chapter, we address the previously outlined requirements and provide an implementation for the introduced concept. As for implementing the key requirements, we chose to mount most of the technology at the workplace instead of requiring workers to wear sensors. Similar to related approaches [14, 99], we use a top-mounted camera-projector system for providing activity recognition and giving in-situ projected feedback. Figure 5.1 shows an illustration of the hardware setup for our assistive system. A top-mounted depth camera is observing the work area and boxes where the parts are stored. Further, a top-mounted projector is able to project in-situ feedback onto the boxes and the work area. Both, depth camera and projector, need to be mounted in a distance so that the boxes and the work area are both covered by the projection area and the field of view of the depth camera.

In the following, the generic parts considering software architecture and implementation are introduced. These parts are designed technology-dependently, i.e. requiring a camera-projector setup.

5.1 Generic Parts

Throughout this thesis, we use a set of data structures and concepts. In this section, we explain the generic parts that are used in our assistive system. These

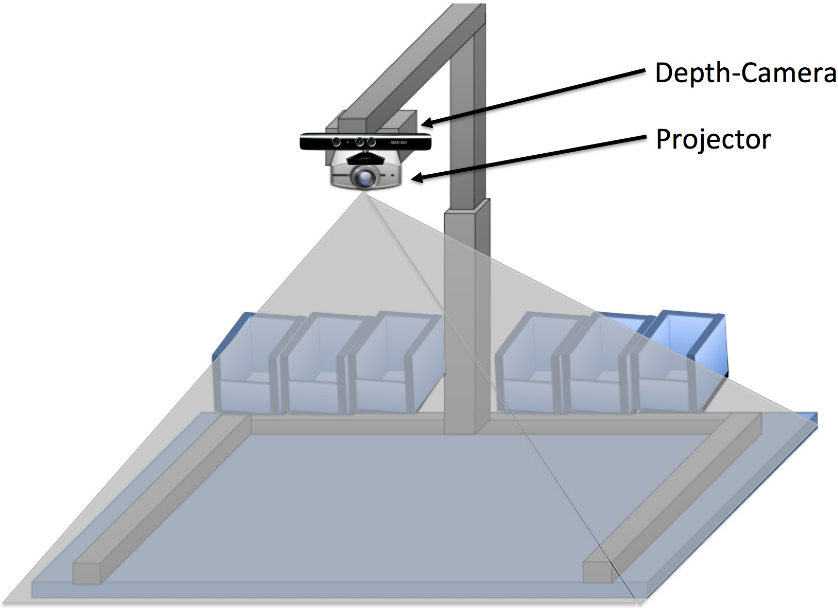


Figure 5.1: An abstract sketch of the technical components that are included in the assistive system for the workplace. It consists of a top-mounted depth camera and a top-mounted projector. The grey area depicts the projection area and the area which is observed by the depth camera.

are in particular: the data structures for workflows, the projected feedback, and triggers for switching between the work steps. Further, we describe the concepts behind the activity recognition and apply the generic parts to different scenarios, which can be found in industrial workplaces.

5.1.1 Workflows

To represent the order of work steps that are performed at the workplace, we define a data structure called *Workflow* to represent a logical sequence of work steps in our assistive system. We defined a *Workflow* as sequence of work steps that contains all work steps performed at the workplace in a defined order. Each logical *Workstep* that is defined in the system represents a physical work step at the workplace. The data structure *Workflow* and its properties are depicted

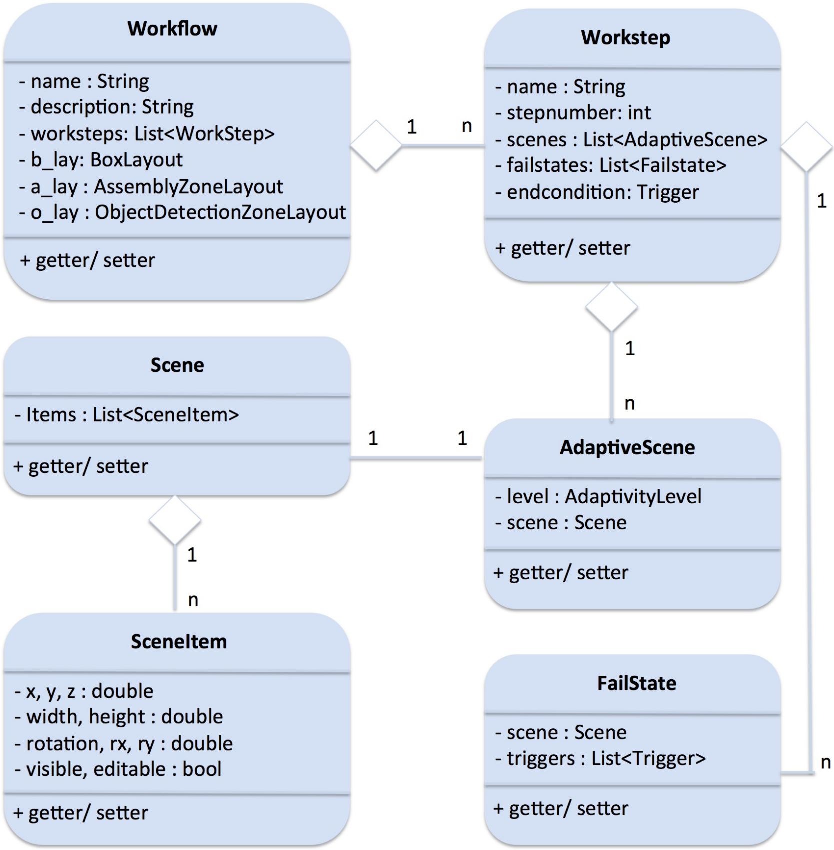


Figure 5.2: The class diagram is showing an overview of the assistive system. The *Workflow* is the central component of the assistive system - it contains all work steps, feedback scenes and *Failstates* that are defined for a *Workflow*.

in Figure 5.2. Accordingly, the *Workflow* is the main data structure that is implemented in our assistive system. It contains a name and a description for its identification. Further, it holds a *BoxLayout*, an *AssemblyZoneLayout*, and an *ObjectDetectionZoneLayout*, which are responsible for activity recognition at the assistive system. They are explained in more detail in the remainder of this chapter. The most important property of a *Workflow* is the list containing one or many *Worksteps*. A *Workstep* is a data structure containing a name for identifying the *Workstep* and a number telling at which position in the defined order the *Workstep* should be performed. Further, it contains a list holding a number of *AdaptiveScenes*, a list holding zero or more of *Failstates*, and a *Trigger* representing the end-condition of the *Workstep*.

5.1.2 Adaptive Feedback

Following the requirement of providing individualized feedback for differently experienced users (**R8**), we implement the concept of adaptive feedback. This is done by implementing the possibility of providing multiple versions of a feedback *Scene* for one work step. Figure 5.2 depicts the class *AdaptiveScene*, which combines a *Scene*, which is a collection of visual feedback during a work step, with an *AdaptivityLevel*. Accordingly, the different scenes stored with the *AdaptivityLevels* contain a different level of detail for providing the feedback. In the assistive system presented in this thesis, we propose using three *AdaptivityLevels* levels: a **beginner level**, an **advanced level**, and an **expert level**.

The design of the three *AdaptivityLevels* follows the idea of successively reducing the amount of displayed feedback with increasing worker experience. In the **beginner level**, the system is providing the full feedback containing text that describes the action, pictures of the correctly performed work step, videos of the work process, and contour information which are projected in-situ - directly at the position where the action has to be performed. In the **advanced level**, the system only displays the contour information to show the position where the action has to be performed. It assumes that the worker is now experienced enough to know the action, however it assists in finding the position faster. When in **expert level**, the system does not provide any feedback because it assumes that the worker is experienced enough to know both work steps and the position where the action has to be performed.

5.1.3 Scenes and Triggers

The most important properties of a *Workstep* are the visual feedback that is shown to the user and the logical conditions for switching between the work steps. In the implementation the visual feedback is represented by a *Scene* and the condition for advancing to the next work step is implemented in a *Trigger*, which are described in this section.

Scenes

A *Scene* is the logical representation of in-situ projected feedback that the user sees while performing a work step. Depending on the *AdaptivityLevel*, only one *Scene* belonging to one *Workstep* is shown at one point in time at the workplace. As depicted in Figure 5.2, a *Scene* can contain many *SceneItems*. A *SceneItem* is a piece of in-situ projected content, e.g. a shape or an image, which is projected at a position (X/Y/Z) at the *Scene*. As the projector can only handle a 2D scene, a position with X and Y coordinates would be sufficient. However, we implemented a vertical Z-position to support overlapping *SceneItems* correctly. Furthermore, a *SceneItem* has a width and a height to define the boundaries of the item. Optionally, it can contain a rotation-point (rX/rY) and a rotation angle to rotate the content of the *SceneItem* around the specified point. Furthermore, each *SceneItem* has boolean flags which can be set. A visible flag can toggle between showing the *SceneItem* and hiding it and an editable flag decides whether a user is able to edit the *SceneItem* or not.

In our implementation, we feature several types of *SceneItems*. The *SceneRect* object represents a rectangle which can change its color. Similarly, the *SceneCircle* object represents a circle which has a radius and can change its color. The *SceneImage* object can display an image that is stored on the hard-disk of the assistive system. Similarly, a *SceneVideo* can display a previously stored video. The video can be played back, paused, and replayed according to the work step. Further, the assistive system is able to display text using *SceneText*. The text contains a font size and a font type that can be set for each text. As the *SceneText* only features one line of text, we provide a *SceneTextViewer*, which is able to display multiple lines of text and implements a scrolling functionality. All properties of the *SceneItems* can be set by the user. To enable a faster item creation, we preset the *SceneItems* with properties that are used the most, e.g. green color is the default for new *SceneRect* and *SceneCircle* objects.

Although there is only one *Scene* stored per *Workstep*, the assistive system is capable of displaying multiple *Scenes* at a time. Thereby, the content of each *Scene* is drawn consecutively on the canvas for each frame. In other words: the system can overlay a *Scene* with one or multiple other *Scenes*. Thereby, the main feedback that is shown during the work steps is stored in the *Scene*, which is stored in the *Workstep*. This mechanism can be used to display error feedback, e.g. when picking a part from a wrong box, or providing debug information, e.g. the current frames per second.

Triggers

A *Trigger* is an event that is raised when a certain logical condition was fulfilled. Our implementation of the assistive system uses *Triggers* to advance from one *Workstep* to the next. Internally, each component that can produce or consume *Triggers* registers to a queue. Thereby each component is able to react to relevant *Triggers*. Table 5.1 provides an overview of the *Triggers* that are implemented in the assistive system. The **assembly trigger** is sent when the measured depth data of the assembly fits the previously recorded depth data that was stored for the assembly. The **pick trigger** is sent, when the worker picks a part from a box. When the system detects that a previously defined object was recognized in a defined area, the system sends an **object trigger**. Similar to the assembly trigger, the **error trigger** indicates that the assembly fits previously defined depth data, however, the depth data was defined as a wrong assembly. Thus, the error trigger and the assembly trigger both indicate a match in the depth data, but the error trigger is used to go into a *Failstate*, whereas the assembly trigger advances the feedback to the next *Workstep*. The system can also handle triggers from external events. These **external triggers** are sent over network and used to communicate between multiple assistive systems. When having two assistive systems that have mutual dependencies, the first system can tell the second system when a work step was finished. Finally, we implement a **manual trigger** that is used to send any type of *Trigger* without having to fulfill the logical condition that is required for the condition. This is mostly used to perform wizard of oz studies or for debugging purposes. An example of a manual trigger is using the keyboard, foodpedals, or the wireless presenter to switch between work steps.

All triggers can be used to advance to the next state in the work flow or issue displaying an error message. Usually assembly triggers, pick triggers, object triggers, and external triggers are used to advance to the next step in the workflow. The error trigger is used to display a *Failstate*, when an error was detected (cf. tri-state assembly logic). The *Failstate* leads to displaying a complete *Scene*, which was created especially for this *Failstate*. When the error trigger is not active anymore - meaning that the error has been resolved, the system returns to the previously displayed *Scene* that shows the feedback according to the current work step.

Trigger Type	Description	Trigger Message
assembly trigger	An assembly fits defined depth data	assembly + <id>
pick trigger	A pick was detected in a box	box + <id>
object trigger	An object was detected in an object zone	obj + <id>
error trigger	A *Failstate* was triggered	err + <id>
external trigger	An external trigger event was received	net + <id>
manual trigger	This can manually send each type of trigger	man + <id>

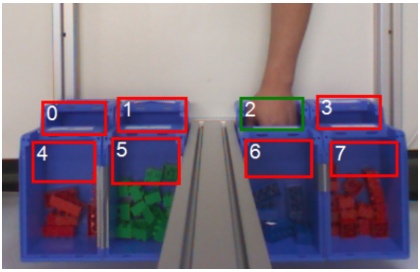
Table 5.1: An overview of the *Triggers* that are used by the assistive system. We implement five categories of triggers: the table lists and describes them and shows the internal trigger message.

5.2 Activity Recognition

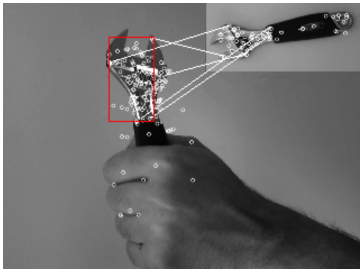
As we choose to equip the workplace with sensors rather than requiring the worker to be equipped with body-worn sensors, we use the RGB image and the depth image to detect activities that are performed at the system. The assistive system is able to perform three types of activity recognition using the RGB image and the depth image: pick detection, assembly detection and object recognition.

5.2.1 Pick Detection

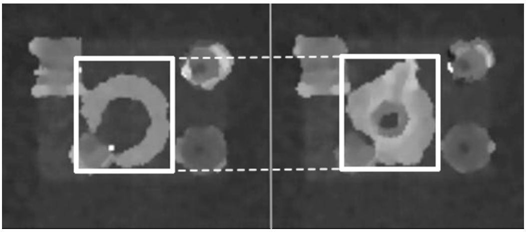
For detecting the picks from bins, we implement an algorithm that surveys the area in front of the bins by observing the depth data in the defined area. The areas in front of the bins are defined by overlaying the bins with a rectangle in the RGB image (see Figure 5.3(a)). As we implemented an accurate mapping between the RGB image and the depth image, we use the RGB image for defining the position of the interactive areas in front of the bins, but use the depth image for detecting the picks. The interactive areas are defined by clicking into the RGB image. They can be adjusted by dragging and dropping the borders of the rectangles that they fit the measurements of the bins. In the implementation of the system, each rectangle that overlays a bin at the workplace is defined as a *Box*, which has a position, height, and width. To detect the picks from the bins, each *Box* stores the mean depth values across all points that are located inside the boundaries of its corresponding rectangle. If a worker picks a part from the bin, each depth point inside the defined rectangle is checked if it is inside the height of the box. If a point is inside the height of the box, it is compared to the mean



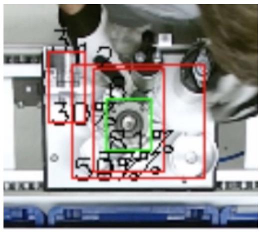
(a) Pick detection based on depth data



(b) Object recognition based on RGB image



(c) Assembly detection based on depth data



(d) Assembly detection UI

Figure 5.3: The components that are used for the activity recognition component in the assistive system. Figure (a) shows the pick detection, Figure (b) shows the tool detection, Figures (c) and (d) show the assembly detection.

depth value that was stored when creating the *Box*. If the difference between the mean depth value and the measured point is over a threshold of 10mm the point is considered as changed. The system sends the pick trigger, when a defined percentage of the points are considered as changed. With our standard picking bin (Bito SK2311: the opening is 11.5cm wide and 5cm high), we found that a percentage of 60% is a robust value to detect most sizes of hands.

Considering the *Trigger*, picking a part from a box results only in sending the pick trigger once. We designed it this way to prevent falsely triggering two consecutive worksteps that might have the same trigger. Therefore, each box implements a timeout that can be set globally for all boxes. Per default, this timeout is initially set to 2000ms. After that, if the worker's hand is still in the bin, the *Trigger* is send a second time, which could lead to a picking error or triggering the next work step by accident. Thus, adjusting the trigger timeout to the picking speed of the worker is necessary.

5.2.2 Assembly Detection

To detect if a part was correctly assembled or not, we implement an assembly detection using the depth data of the top-mounted camera. Therefore, in an initial calibration step, the system has to store the depth data of the correctly assembled part. Similar to defining the zones for the pick detection, the user can use the mouse to draw a rectangle in the RGB image around the area where the part was assembled (see Figure 5.3(d)). This rectangle is internally stored as an *AssemblyZone*. The position, depth data, and size of multiple *AssemblyZones* are stored in an *AssemblyZoneLayout*, which are properties of the workflow (see Figure 5.2). Upon creating an *AssemblyZone*, the system stores the depth data of the correctly assembled part as a property. Once the correct assembly was stored, the system compares the depth data inside the observed area 30 times per second to the previously stored depth data of the correctly assembled part, which is stored in the *AssemblyZone*.

The assembly detection algorithm works by performing a pixel-wise comparison on the depth values. Thereby, each depth value is compared to the depth value that was stored at the corresponding position. A depth value is considered as matching the previously trained depth value, if it is within a range of $+3$mm and -3mm. If at least 85% of the depth values inside the rectangle are considered matching, the system triggers a correct assembly of the part (see Figure 5.3(d) - green rectangle). Otherwise the system shows a red rectangle. The assembly trigger is activated whenever a correct assembly was detected. In case of a correct assembly this trigger can be sent up to 30 times per second.

The threshold of 85% was determined empirically and provides a good accuracy for our experimental setup. When choosing a higher threshold, a better accuracy can be achieved. However, if the threshold is too high, a correctly assembled part might not be detected anymore because of sensor noise. To correct for the sensor noise, both capturing the reference image and comparing it to the current depth data in the assembly is done on a smoothed depth image. Therefore, we use the last 15 frames of the depth image and calculate the mean value. By doing this, outliers are corrected and a detection can be done more robustly. Using our lab setup, where the camera is 96cm above the assembly area, we are able to robustly detect the assembly of parts that have the size of 1cm$\times$1cm$\times$1cm. Parts that are smaller, cannot be detected accurately using a Kinect_v2 due to its resolution, distance to the assembly area, and sensor noise. In other installations e.g. the assembly cell, we are using a threshold of 80% for correct assembly. In this setup, the Kinect_v2 was 1.6m above the assembly area, which results in the minimum part size that can be detected being bigger than 1cm$\times$1cm$\times$1cm.

5.2.3 Object Recognition

Further the assistive system is providing an object recognition component that is capable of detecting the tools that are used in a workflow. As in the industry, tools that are used during an assembly process have a defined position, we designed the object recognition component to work in defined parts of the camera image - the so called *ObjectDetectionZones*. By using only a part of the camera image for the object recognition, the system saves processing time. Multiple *ObjectDetectionZones* can be defined and stored in an *ObjectDetectionLayout*, which is stored as a property in each workflow (as described in Figure 5.2). An *ObjectDetectionZone* is defined similarly to the *AssemblyZone* and the *Box* - just by using the mouse to define a rectangle around the area where the system should detect the presence of an object.

Inspired by previous work [62], we designed the object recognition component to have an initial calibration step, where the user can place the object that should be recognized inside the *ObjectDetectionZones*. Once the object is inside, the user can press a button, which takes an RGB picture of the object. By doing this calibration step, the reference image has the same angle and distance to the *ObjectDetectionZone* than using it while working at the assistive system, which will decrease the needed computing power (compared to a high resolution reference image). Once a reference image for an object was defined, the system uses the Speeded Up Robust Feature (SURF) algorithm [17] to compare the features of the reference image to the features of the RGB image. If the SURF algorithm finds five features that are both in the reference image and in the current RGB image, the system considers the object as found and sends an object trigger. To save computing resources, we limited the number of performing the object detection to three times per second. With the 30 frames per second that the system gets from the Kinect_v2, this results in using every 10th frame. This means that if an object is present at the *ObjectDetectionZone*, the object trigger is sent a maximum of three times per second. As the used SURF algorithm is rotation-invariant and scale-invariant, it can detect the tool inside the *ObjectDetectionZone*, in each distance or angle. This makes the SURF algorithm a perfect choice for detecting tools that are put on a defined area at the workplace.

5.3 Feedback through In-Situ Projection

For reacting upon the actions that are detected in the activity recognition component, the feedback component uses in-situ projection as a central method of giving instructions. Therefore, the top-mounted projector is used to highlight the correct picking bins and the correct assembly positions. Throughout this thesis, we agree on using the following color concept for giving feedback: To indicate the next picking bin, the system highlights the bin using a green light. For reacting upon picking a part from a wrong box, the system highlights the wrong box that is picked from using a red light. Further, for communicating the assembly position of a picked part, the system highlights the position on the assembly area using a green light. Sometimes parts have to be removed from the assembly area. To indicate the worker that a part should be removed, the part is blinking red. The feedback is provided by displaying the previously explained *Scenes* on the top-mounted projector.

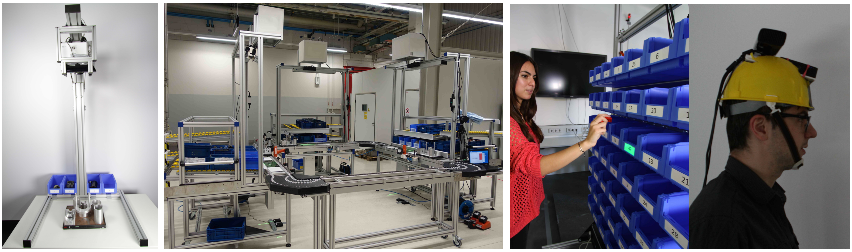
5.4 Applying the Concept in Different Scenarios

As we defined the requirements for assistive systems considering multiple scenarios in industrial applications, and we designed the concept as generic as possible, we are now able to apply the previously introduced assistive system implementation to four prototypes in two scenarios that can be found in industrial manufacturing: a manual assembly workplace, an assembly cell workplace, and two prototypes for an order picking scenario.

5.4.1 Manual Assembly Workplace

The first scenario is the manual assembly workplace, where one worker is working autonomously at one workplace and performs a defined number of work steps. The manual assembly workplace does not have direct dependencies on other workplaces, as work pieces are either started to be assembled at this workplace or a large bin is used as a buffer for the parts that are required for the assembly.

In our lab setup for the manual assembly workplace, we are using an aluminum frame that is holding a Kinect_v2 and an ASUS K330 projector at 96cm above



(a) Single assembly workplace (b) Assembly cell with multiple workplaces (c) Order picking prototypes

Figure 5.4: The concept is applied to four prototypes in two different scenarios: (a) at a manual assembly workplace with one worker, (b) at an assembly workplace for multiple workers that are dependent on each other, and (c) to give cognitive support during order picking tasks using a user-mounted and a cart-mounted prototype.

the work area. Figure 5.4(a) shows the lab setup of the system. Throughout the work, we apply the manual assembly workplace to different setups according to manufactured products used in the different studies. The setups used in the studies have different settings which are tailored especially to the products that are used.

5.4.2 Assembly Cell

The second prototype for the workplace scenario is the assembly cell. An assembly cell is a concatenation of manual assembly workplaces that are dependent on their predecessor. In our setup the assembly cell is designed in a U-shape and uses multiple workpiece carriers that are transported in a circle using a roller conveyor.

We are using this setup to assemble a car's engine starter. In this setup, we were using three pairs of a Kinect_v2 and a Casio AJ-X251 projector. The projector and the Kinect_v2 were mounted approximately 160cm above the assembly area. Figure 5.4(b) shows the assembly cell in an industrial setup. The work steps that are performed at the assembly cell were balanced were balanced in a way so that it takes the equal amount of time in each of the three assembly workplaces to assemble the product. This balancing is necessary that the workplaces do not have to wait for each other.

5.4.3 Order Picking

As a second scenario, we apply the concept of an assistive system for the workplace to provide cognitive support while performing an order picking task. Therefore, we created two assistive systems for order picking: an augmented picking cart and an augmented helmet. The picking cart carries three pairs of a Kinect_v1 and a projector (see Figure 5.4(c)) and the augmented helmet carries one pair of Kinect_v1 and a projector. In these two systems, only the pick detection is used as activity recognition component for detecting the picks from boxes and the compartments. Further, the order picking cart and the augmented helmet can project in-situ feedback onto the environment. The detailed description of the order picking cart and the helmet can be found in Chapter 7.

INFO: As a part of the motionEAP project, we also applied our assistive system to a stationary order picking system. This project is described in detail in the PhD thesis of Andreas Bächler.

Chapter 6

System: An Augmented Workplace

This chapter introduces the assistive system that is used in the manual assembly workplace scenario and in the assembly cell scenario. Using the hardware setup and the general software architecture that was introduced in the previous two chapters, this chapter focuses on the user interface of the software and the possibilities for interacting with the system for both engineers and workers. The basis for this chapter is our experimental lab setup (see Figure 6.1), however the same user interface and all functionality is also available in the assembly cell setup.

The lab setup consists of a Kinect_v2 and an ASUS K330 projector that are mounted on top of the aluminum construction in a way that they are both facing downwards (see Figure 6.1 (A)). Further, a variable amount of picking bins are placed at the back of the assembly area - we call this the picking area (see Figure 6.1 B)). The picking bins can be stacked on top of each other, as the pick detection also works when there is only a difference in Z-direction between two picking bins. Also considering the in-situ projection stacking the picking bins is feasible as the projector is mounted with an angle in a way that the projection will not be occluded by the top-most picking bin. For the lab setup, we are using the Bito SK2311[12] picking bins. On the bottom of the assistive

[12] Bito SK2311 - `http://www.bito.com/Artikel/Sichtlagerkaesten-SK-1449.html?CatalogCategoryID=rRzAqAGOy7IAAAE1dCoYgGIY`

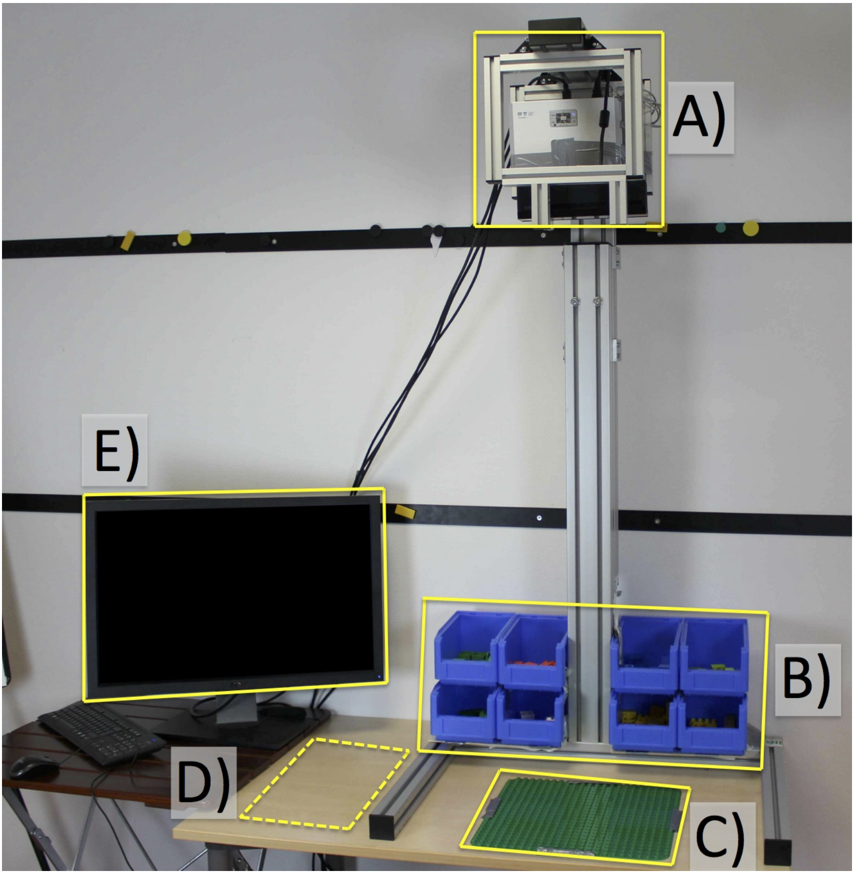


Figure 6.1: The lab setup of our assistive system. The screen next to the assistive system is from the experimental setup that is described in Chapter 8.(A) The top-mounted projector and depth camera. (B) The bins where parts for the assembly are stored. (C) The assembly area where assembly steps are performed. (D) The tool area where tools are placed and recognized by the system. (E) A nearby monitor that is used by the engineer to setup the system.

system, there is a defined area for assembling the parts. We call this the assembly area (see Figure 6.1 (C)). The system uses the depth data that is recorded in the assembly area to check for correct assembly using the assembly detection algorithm. Usually, at the assembly area there is a workpiece carrier holding the assembly in a position that the assembly steps can be detected by the top-mounted Kinect_v2. In the depicted setup, we use a Lego Duplo plate as a workpiece carrier for an abstract Lego Duplo assembly task. Finally, the area next to the assembly area is the tool area (see Figure 6.1 (D)). The tool area can be defined on any free surface on both sides of the workplace. As the RGB image that the Kinect_v2 is recording has a wider field of view than the Kinect_v2's depth image and the object recognition algorithm only uses RGB data, the system can be calibrated so that the depth image only covers the assembly area and the picking bins. Finally, the setup uses a nearby monitor to configure the system (see Figure 6.1 (E)).

6.1 User Roles

As not every feature in the assistive system is relevant for each user, we grouped the features of the assistive system according to two different user roles: the engineer and the worker. In the following the two user roles, their tasks, and their way of interacting with the system are defined.

The engineer is setting up the system, adjusting the parameters to the environment and the task, and creating new workflows. Setting up the system includes performing the camera-projector calibration, defining the assembly zones, defining the interactive boxes that are created above the picking bins, and defining the tool areas. For creating workflows, the engineer is able to use the manual workflow editor mode where the engineer can drag and drop interactive items directly onto the work area. Further the engineer can use the Programming by Demonstration mode, where workflows are defined by performing the assembly once. As a result, the engineer has to adjust all available parameters of each activity recognition module according to the assembled product and the production environment.

The worker only uses the graphical user interface that is displayed on the monitor for checking the progress of the workflow. However, this visualization of the work progress can easily be displayed directly on the work area. Every other interaction with the system is done implicitly by performing work steps directly at the workplace. The system is sensing the worker's actions and is reacting to the actions using the three activity recognition modules: pick detection, assembly detection, and tool detection (cf. Section 5.2).

6.2 User Interface and Interaction Design

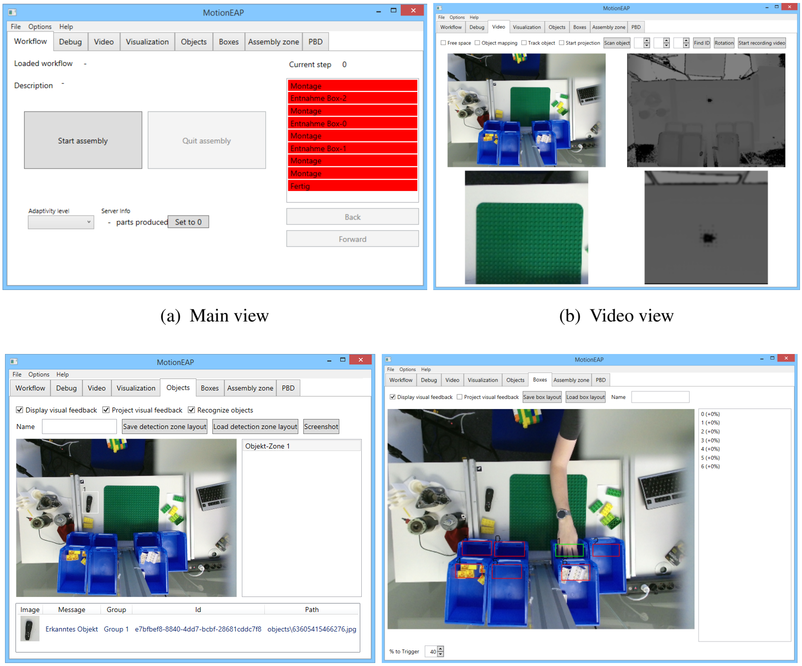
Considering the user interface, we group the interaction with the system according to the user role. As engineers use different functions for installing and setting up the assistive system, we start describing the engineer's user interface first and describe the worker's interface afterwards.

6.2.1 Engineer User Interface

As the engineer's main task is to setup the assistive system at the workplace and adjusting it to the given conditions, the engineer has access to more functionality than the worker. This functionality is accessed using the graphical user interface that is displayed at the monitor next to the assistive system (see Figure 6.1 (E)). The different views are either tabs that are contained in the main window, or pop-up dialogs where the engineer can change global settings or perform calibration steps.

For getting an overview about the loaded workflow, the engineer can look at the *main view* (see Figure 6.2(a)). Here the loaded workflow is displayed with a description, a name, and all work steps which are included in the workflow. The user interface provides buttons to jump between the work steps without having to fulfill their end conditions. This is mostly used for debugging purposes by the engineer. Further, the *main view* provides information about the currently active adaptivity level (cf. Section 5.1.2). Additionally, there is a counter that displays the number of produced parts that were produced so far.

One of the engineer's tasks is to align the camera according to the work area. Therefore, the assistive system provides a *video view* where four different live images are presented (see Figure 6.2(b)). The upper left video is the RGB video of the camera's whole field of view that contains both depth and RGB data. This is mainly used to center the camera at the workplace and to quickly see if every important component is seen by the camera. The upper right video is the depth data of the camera's whole field of view. The lower left video is an RGB video that only depicts the assembly area. The assembly area has to be defined as only in the assembly area the CPU-intense operations for the assembly detection are performed. This video is mainly used by the engineer to inspect if the whole workpiece carrier is covered by the assembly area. Finally, the lower right video is a smoothed version of the depth data of the assembly area. Thereby, the system calculates the mean depth data using the 15 previous frames. This smoothed



(a) Main view (b) Video view

(c) Objects view (d) Boxes view

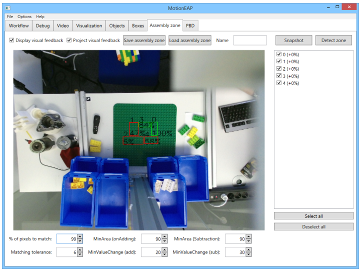
Figure 6.2: An overview of the user interface that can be accessed by the user groups using the monitor. (a) The main view showing the current work step and loaded workflow. (b) The video view showing the RGB and depth view from above. (c) The objects view showing the known objects. (d) The boxes view showing the interactive picking areas.

depth data is also used for the assembly detection, as operating on live depth data without filtering would lead an inaccurate assembly detection. In case there are multiple cameras connected to the system, the engineer uses the video view to choose a camera that should be used for the pick detection, the assembly detection, and the object recognition.

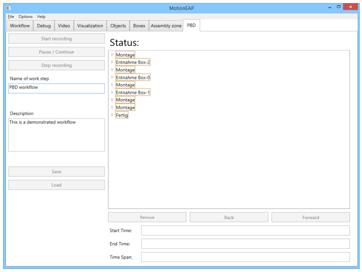
For enabling the object recognition for workers, the engineer first has to define object detection zones in the *object view*. To create a zone, the engineer clicks at a point in the camera image. All object detection zones are listed on the right side of the view. Once the zone is set up and the zone is marked for recognition, the system will constantly try to find known objects in the zone. The system also contains a database of known objects, which are shown in the list at the bottom of the object view. The view also contains several settings related to object recognition. The engineer can e.g. display visual feedback using the projector to better see the borders of the object recognition zones, or display the borders of the zones directly on the camera image. Further, this view can save the layouts to a file and load them again. Finally, this view can be used to register new objects with the database by pressing the screenshot button.

Another task of the engineer is to define the interactive boxes, which trigger when a worker is picking a piece from a picking bin. The engineer can define these interactive boxes using the *boxes view* (see Figure 6.2(d)). For creating a new interactive box, the engineer clicks at the position where the new box should be created in the camera image of the boxes view. The system will create a new box at the defined area using the click position as the upper left point of the box. It creates a standard box with pre-defined size and the Z-position of the click position as a reference value. All created boxes are displayed in the list on the right side of the boxes view. To delete a box, the engineer can right click on the box in the list and select the delete option from the drop-down menu. To adjust the width and height of the box, the engineer can either use the borders that are drawn in the camera image of the boxes view, or use a textual interface that can be accessed by right-clicking a box in the list and selecting properties. In this properties dialog, the engineer can change parameters of the box e.g. trigger condition, height, width, or Z-position. On the boxes view itself, the engineer can adjust global parameters, e.g. the trigger sensitivity which is represented by the percentage of pixels that need to change for the system to trigger a pick.

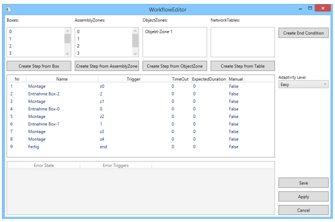
The *assembly view* (see Figure 6.3(a)) is used to create interactive assembly zones for checking for correct assembly of work pieces. To create an assembly zone, the engineer first assembles the work piece at the correct position. Then, the engineer has to define where the assembly takes place. This is done by clicking into the camera image at the assembly position. The assembly zone can be adjusted by dragging the borders of the assembly zone until it encloses the complete assembly. Alternatively, assembly zones can also be created automatically by taking a screenshot before and after the assembly. The system then tries to find the assembled piece automatically based on the change in the depth data. After creating the assembly zone, the system constantly checks if the depth data inside the zone matches the correctly assembled depth data that was stored when creating the zone. Multiple assembly zones can be created using the assembly view. The created assembly zones are shown in the list on the right side of the view. The engineer can store and load created assembly zone layouts using the save and



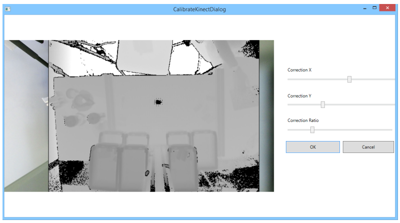
(a) Assembly view



(b) Programming by Demonstration view



(c) Edit workflow view



(d) Calibrate Kinect view

Figure 6.3: An overview about the user interface that can be accessed by the user groups using the monitor. (a) The assembly view which shows the assembly zones that detect correct assembly of a work piece. (b) The Programming by Demonstration view which allows teaching new workflows by demonstrating them. (c) The workflow editor, which allows creating new worksteps and modifying them. (d) The Kinect has to be calibrated manually to combine RGB and depth data using the calibration view.

load button on the assembly view. Assembly zones can be deleted by opening the context menu on the list of assembly zones and selecting the delete option. The parameters e.g. trigger, height, width, and Z-position can be changed in the properties menu that can be accessed by the engineer by opening the context menu. The settings that can be changed globally are done directly in the assembly view at the bottom. The engineer can set the percentage that has to match the depth data in order to detect a correct assembly or the tolerance which considers depth data as correct or not.

For creating a workflow, the engineer has two options: Using the *Programming by Demonstration view* to demonstrate work steps to the system or using the *edit workflow view* to create the work steps manually. The *Programming by*

Demonstration view is depicted in Figure 6.3(b). The engineer can control the recording of the work steps by pressing the buttons start recording, pause/continue recording, and stop recording. Once the recording is on, the engineer can pick and assemble parts according to the workflow that needs to be teached. For every action the engineer makes, a work step is created in the list of work steps on the right side of the view. If the engineer makes a mistake, work steps can be removed again using the button at the bottom of the view. Once the recording is done, the workflow can be saved to a file. Later, the workflow can be loaded again to further adjust or graphically enhance the workflow.

For creating a workflow manually from different triggers or to alter existing workflows, the engineer can use the *edit workflow view* (see Figure 6.3(c)). There, the engineer can create work steps with trigger conditions according to existing boxes, assembly zones, object detection zones, or external triggers. The triggers that are known to the system can be selected from the four lists on top of the edit workflow view. An end condition can be created by using the create end condition button. The order of the work steps can be adjusted in the list in the middle of the view by dragging and dropping the work steps at the desired position. Once a work step is selected, the system will project the Scene that is shown during the selected workstep on the projector. The engineer can use this view to drag and drop SceneItems onto the projection area and adjust them (see an earlier version of this view in Figure 10.4(a)). To switch between the different adaptivity levels, the engineer can select the current adaptivity level on the right side of the view. To enable a faster creation of workflows, work steps can be copied and pasted. For creating error states, the engineer can select triggers that should be treated as an error and drag them into the bottom list of the view. For saving workflows to a file or for applying them directly to the system, the engineer can press the save or apply button at the lower right side of the edit workflow view.

As each installation of the assistive system might have a different camera position and a different hardware setup, the engineer has the opportunity to calibrate the depth and RGB image of the Kinect_v2 according to the workplace using the *calibrate kinect view* (see Figure 6.3(d)). Here the engineer can correct the depth image in X and Y direction and can set a ratio, which stretches the depth images to fit the RGB image.

6.2.2 Worker User Interface

The worker's main task is to use the assistive system implicitly by performing the work tasks that are required for the assembly. Therefore, all actions that the worker can perform are in the work area and not on the GUI. Figure 6.1 shows a single workplace and all its components and interactive areas. Before the worker is starting his or her work task, the engineer is starting the workflow using the GUI. Then, the worker can start working. The worker can trigger picks by picking parts from the picking bins (see Figure 6.1 (B)) and can trigger assembly steps by correctly assembling picked parts at the work area (see Figure 6.1 (C)). Additionally, the worker can use tools that are stored in the tool area (see Figure 6.1 (D)). To check the produced number of parts and to check the current step in the workflow, the worker can look at the nearby monitor, which is depicted in Figure 6.1 (E).

6.3 Feedback Adaptivity Levels

We implemented our assistive system to support three different adaptivity levels: beginner level, advanced level, and expert level. In the following, we define which type of feedback is presented in each level. Figure 6.4 shows a workplace for assembling a car's engine starter in the advanced level. As this is an instruction concept that resulted from the studies conducted in this thesis, the feedback that was provided during the conducted studies differed from this concept.

Beginner Level
In our instruction concept, in beginner level we provide full feedback including video, contour, and textual information. Video instructions are shown at a designated video area. The video instructions are recorded from the worker's point of view. In case of a picking step, no video is shown as an instruction. As an additional information where to assemble or pick a part, we show a contour information that illuminates the position and orientation of the part to assemble in an assembly step. During a picking step, the bin to pick from is highlighted using a green rectangle in front of the bin to pick from. In beginner level, the action that needs to be performed next is also displayed using a text. For identifying the picking bins, we also project a description of the parts in front of the picking bins. We further integrated a progress bar showing the current progress to show the worker, where in the workflow the assembly is at the moment.

Advanced Level
In the advanced level, the feedback that is shown is reduced to not overload the worker with information. Therefore, we only show the contour information and the textual information in the advanced level. To keep consistent with the

Figure 6.4: The projected feedback can be enhanced with additional information. In the beginner level, a video is shown additionally to text, image, and contour feedback. In advanced level only the contour is shown. In expert level, no visual feedback is shown at all.

beginner level, in the advanced level we also display the progress bar and the description of the picking bins.

Expert Level
For experienced workers, we implemented the expert level. Here only the description of the parts and the progress bar is shown to the worker for indicating that the system is still running. As the system's activity recognition components are not turned off in expert level, the system still can display an error message in case the worker performs a work step incorrectly.

6.4 Switching the Adaptivity Level

With our assistive system, we introduce a concept for switching between the three previously introduced adaptivity levels. As a basic way of determining the appropriate adaptivity level, we consider the error ratio in the last 100 performed work steps. If a worker would make one error out of 100 performed work steps, the error ratio would be 0.01. If there were not 100 work steps performed yet, the system starts in the beginner level. Once an error ratio out of 100 work steps can be calculated, the system switches the adaptivity level according to the following thresholds: If the error ratio is greater than 0.05, then the system presents the instructions in beginner level. If the error ration is between 0.05 and 0.01 then the system presents the instructions in advanced level. Finally, if the error ratio is below 0.01, the system switches to the expert level.

Apart from the initial adaptivity level classification, the system can change the adaptivity level temporarily according to two events while a work step is performed. The first event is when an error was made. This could be a picking error or an assembly error. In that case, the adaptivity level is immediately decreased temporarily for the current step. In case the error ratio became larger than a threshold for switching the level resulting from the currently made error, the next work step is presented in the adaptivity level that matches the current error ratio. The second event that temporarily decreases the adaptivity level is when the user is taking too long for the work step. As the system provides an option for specifying a maximum time for each work step, the system uses this time to assume if the worker is having a problem with the current work step. After the maximum time for the work step has elapsed, the system decreases the adaptivity level immediately and displays more instructions for the current work step.

As we tried this concept in a long-term industry setting, we suggest to further use physiological data to determine the appropriate adaptivity level for the worker. If the system would have access to live physiological data, the system could detect, which activities are causing stress for the worker and immediately reacting upon stress by changing the adaptivity levels [48].

6.5 Vision: A Product

We expect that augmenting workplaces with assistive systems will have a huge impact on manual assembly workplaces in the future. As the system that is presented in this thesis is a research prototype for evaluating the general concepts behind using projection-based Augmented Reality (AR) at workplaces, we envision a more compact system as a product for the future. Similar to the anglepoise lamp proposed by Linder et al. [107], we think that the form factor of a lamp might be beneficial for assistive systems using projection-based AR at manual assembly workplaces. As state-of-the-art workplaces are equipped with a work-

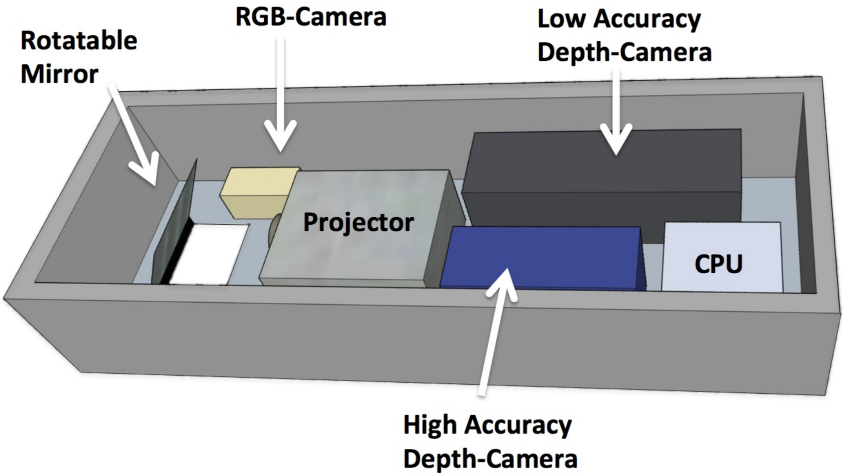


Figure 6.5: We envision that every component can be placed in a single lamp that can be placed over the workplace.

place luminaire, the form factor of a workplace luminaire might be convenient. Figure 6.5 shows a conceptual 3D drawing of a workplace luminaire equipped with the technology that is necessary for an assistive system. In our concept, we integrate a small projector into the workplace luminaire and use a mirror to steer the projection at the desired area on the workplace. A small RGB camera is also integrated to take pictures for quality control and for enabling object recognition algorithms. In our concept, we integrate two depth cameras. A low accuracy depth camera with a wide field of view for detecting the worker's hands and detecting the picks from the picking bins and a high accuracy depth camera for assembly detection. Our tests indicate that a low accuracy depth camera is very suitable for pick detection. However a problem with low accuracy depth cameras is the assembly detection. In our tests, the Kinect_v2 still introduces noise in the depth data that makes robustly detecting small parts (e.g. screws or washers) impossible. Thus, we experiment with high accuracy depth cameras for the assembly detection. In our envisioned product, we are using an Ensenso N10[13] depth camera for the assembly detection and an Asus Xtion for the pick detection. The computation is done on a small CPU unit (e.g. a LattePanda[14]) which is also integrated directly in the workplace luminaire.

[13] Ensenso N10 - `http://www.ensenso.com/portfolio-item/n10` last access 5th Oct. 2016

[14] LattePanda - `http://www.lattepanda.com/` last access 5th Oct. 2016

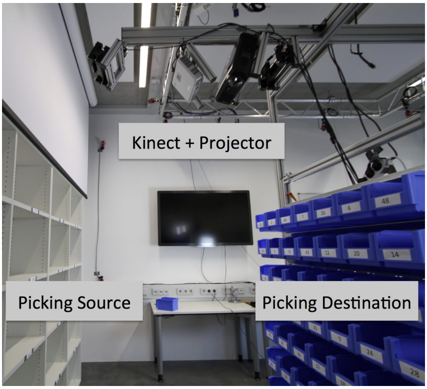
Chapter 7

An Order Picking Support System

This chapter is based on the following publications:

- M. Funk, A. S. Shirazi, S. Mayer, L. Lischke, and A. Schmidt. Pick from here!: An interactive mobile cart using in-situ projection for order picking. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 601–609. ACM, 2015
- M. Funk, S. Mayer, M. Nistor, and A. Schmidt. Mobile in-situ pick-by-vision: Order picking support using a projector helmet. In *Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, New York, NY, USA, 2016. ACM

In this chapter, we introduce two systems to provide cognitive support for workers during order picking tasks: a cart-mounted system holding three camera projector pairs (OrderPickAR) and a head-mounted system, where the user carries a camera-projector pair, which is mounted on a helmet (HelmetPickAR). Both systems adapt the implicit interaction concept and the requirements that were introduced in



(a) Components of the order picking cart

(b) A compartment is illuminated by the cart

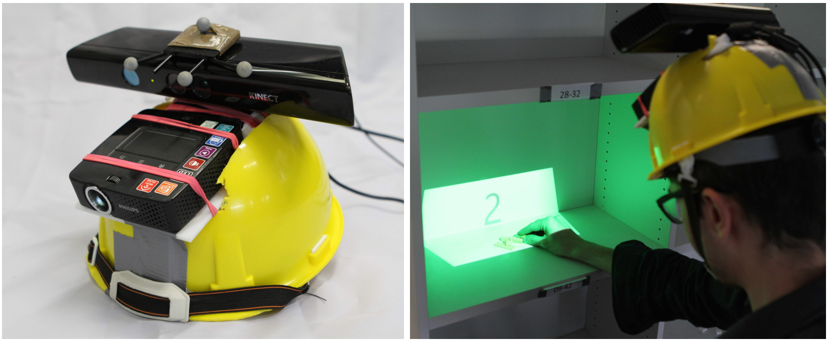
Figure 7.1: The OrderPickAR mobile cart for order picking. (a) An overview of the components used in the system. (b) The system during use: The cart illuminates the compartment the worker has to pick from next and projects the amount of items to pick.

Chapter 4. They further implement the general software parts that were introduced in Chapter 5.

7.1 System: An Order Picking Cart

In the cart-mounted prototype, which is called OrderPickAR, we augment a regular order picking cart for a classical man-to-goods system with a top-mounted beam holding three pairs of camera-projector (see Figure 7.1(a)). Two pairs are facing the bins of the cart holding the processed orders located at each side of the cart. One pair is facing the compartments of the warehouse containing the items that can be picked. The cameras mounted on the cart are Kinect_v1 depth cameras that monitor the bins mounted on the cart and the shelves in the warehouse. The projectors are used for providing in-situ feedback by highlighting compartments to pick items from, and bins to store the processed orders.

The Field of View (FoV) of the cart-facing camera-projector pairs cover all bins that are mounted on the order picking cart. As they are mounted firmly on the top-mounted beam, they are moved together with the bins that are mounted on the



(a) The projector helmet prototype (b) User wearing the system

Figure 7.2: The projector helmet prototype (a) consists of a Kinect_v1 and an Android projector. (b) The system highlights the location of the parts to pick and displays the quantity.

cart. Thus, the layout of the bins can be predefined and stays constant even when moving the cart. In our prototype, the cart holds 49 bins on each side (a 7×7 grid) and can store up to 98 orders at the same time. Furthermore, the height of the cart can be adjusted to the warehouse, as we constructed the frame holding the beam to be height-adjustable. In our configuration, the height of the cart was set to 3.38m to perfectly cover the bins and the height of the shelves.

The cart is built from aluminum profiles which are typically used in the industry. The projectors facing the cart are ACER K335 LED-projectors with 1000 ANSI Lumen. The projector facing the shelves is an Optoma EW610ST DLP projector with 3100 ANSI Lumen. The depth cameras are Kinect_v1 for Windows running with a 640×480 depth resolution. At the bottom of the cart, we installed a PC that runs the pick detection and calculates the projection on both sides of the cart and in the environment. The system is powered through a ceiling-mounted electric wire.

7.2 System: A Projector Helmet

Inspired by Schwerdtfeger et al. [142], for the head-mounted prototype we chose to augment a helmet with a Kinect_v1 and an small projector. We call this system HelmetPickAR. The HelmetPickAR system is projecting picking information directly into the FoV of the worker. As the system is user-worn, it is mobile and can be used with any type of order picking warehouse. The system is capable of navigating the worker to the next compartment to pick from and projecting the number of items to pick directly into the compartment. When the worker is standing in front of the compartments, the FoV of the projector is smaller than the width of the compartment (see Figure 7.2(b)), which is 57cm×37.5cm×30cm. Thus, the system can only highlight the compartment to pick from when the worker is standing directly in front of it. To guide the worker to the correct picking position even when the worker is not looking in the direction of the next compartment, HelmetPickAR can show arrows that are pointing towards the direction of the next compartment to pick from.

The HelmetPickAR prototype consists of a standard building-site helmet where we cut out the top (see Figure 7.2(a)) to insert a plastic plate carrying the projector and the Kinect_v1. For the prototype we are using a Philips Picopix PPX3610 projector with 100 ANSI Lumen that is connected to a laptop using HDMI. We chose to use wires that connect the helmet to the laptop transferring the depth image of the Kinect_v1, the video feed of the projector, and powering both Kinect_v1 and projector using an external power supply. The laptop is responsible for calculating the instruction that is shown to the user, depending on the position of the user, the orientation and angle of the helmet, and the current picking task according to the workflow.

To enable the workers to interact with the system, we implemented a user interface using an interactive floor display. The floor display is activated when the worker holds the head in an angle of 15° or lower. Once the floor display is activated, it displays four interactive zones where the worker can control the workflow. We considered implementing the following four actions: previous workstep, next workstep, restart the workflow, and end the workflow. For implementing the floor display we are using the Kinect_v1's built-in inclinometer. Once the worker's head is in an angle of 15° or lower, the system waits until the head is at a stable position (no changes in the angle within a threshold of 5° and not more than 10cm difference in the height between the left and right border of the depth image), captures a reference depth image of the floor, and creates four interactive areas in the depth image. The floor display uses the same algorithm for the interactive floor areas than detecting picks from bins (cf. Section 5.2). If the user now touches an interactive area with a foot, the area is activated and a trigger is send to the system. As the floor display is very sensitive to when the user moves the head, the system deactivates the interactive areas when the worker's head moved more than 5° after recording the reference image or if more than two interactive areas trigger simultaneously. The system then waits for the user to hold the head in a stable position again and captures another reference image. To prevent the projector from glaring other workers, HelmetPickAR uses the position of the OptiTrack system to only activate the projection, when the user is facing the compartments to pick from, the destination of the picked items, or holding the head in an angle of 15° to activate the floor display for interaction.

7.3 Making the Systems Step-Aware

To achieve that the systems become step-aware, they have to detect when a worker is picking an item from a compartment and when the worker is placing a picked item at the destination bin. Both systems use the Kinect_v1 to detect the worker's hands entering previously defined zones. Additionally, both systems require knowledge about the 3D model of the warehouse, i.e. the position of each compartment to pick from and the position of each target bin to place the picked items. In the OrderPickAR system, this warehouse layout can be created using a graphical editor (see Figure 7.3(a)). In contrast, HelmetPickAR uses an XML-file holding the position, width, and height of the compartments and target bins of the warehouse layout. The graphical editor is developed using Unity3D[15]. Using the editor, the user can define interactive zones, so-called trigger spheres, to overlay the compartments of the shelves in the model at the position where the compartments are located in the physical world. In Figure 7.3(a) the trigger spheres that are shown in (1) are the spheres that are responsible for detecting the picks from the compartments. The spheres that are shown in (2) are the spheres that are responsible for detecting if a picked part was placed in the correct target bin. For the OrderPickAR prototype, the target spheres are moving together with the position of the cart. However, in the scenario that is used for HelmetPickAR, the target bins are located at a fixed position in the warehouse.

To track the position and orientation of both cart and helmet in the warehouse, we use the OptiTrack[16] motion capturing system. We equipped the warehouse with 17 OptiTrack Flex3 cameras and positioned a marker at the upper frame of the cart, and on top of the helmet to detect the systems' orientation and position in the warehouse. We further tracked the beginning of the warehouse using a marker instead of defining the shelves at a fixed position in the 3D model. The OptiTrack cameras were positioned throughout the warehouse in a way that for

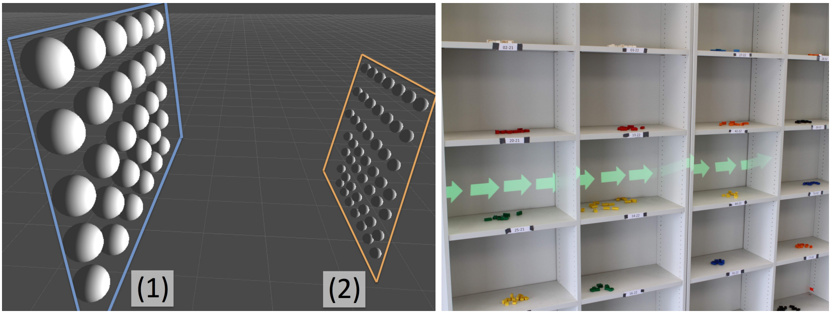
[15] `www.unity3d.com` (last access 5th Oct. 2016)

[16] `www.naturalpoint.com/optitrack` (last access 5th Oct. 2016)

every possible position of the systems, at least four cameras were able to track the markers. According to the specification, this allows the OptiTrack system to track the position and orientation of the markers within an accuracy of millimeters. The OptiTrack system is connected to a dedicated computer for handling the OptiTrack cameras and calculating the position information of the markers. The position and orientation information is then streamed to the computer of the cart and the computer of the helmet via WiFi.

For detecting the picks in the warehouse, the algorithms in both prototypes follow the same principle as the pick detection described in Section 5.2. The systems align the depth image that is taken by the Kinect_v1 with the position and orientation that is received from the OptiTrack system and use the data to observe the trigger spheres. They calculate the average depth value inside the trigger spheres and compare it to the previous state. If the change in depth data is beyond a threshold, the sphere triggers that the user picked from the associated compartment. Using the cart, an informal experiment suggested using a threshold of 61% concerning the changed depth pixels for reliably triggering the interaction for our compartments. For HelmetPickAR, the threshold needs to be adjusted to the worker. The threshold value is also dependent on the size of the compartment, as the interactive area changes according to the compartment's size.

For detecting when a worker places a previously picked item, OrderPickAR and HelmetPickAR use different approaches. OrderPickAR uses its cart-facing depth cameras that are mounted at the cart's beam in a 90° angle. They are used like a light barrier and can detect when a user is putting an object or a hand into one of the cart's bins. Therefore the cart uses mobile trigger spheres that are depicted in Figure 7.3(a) (2). As the position of the trigger spheres move according to the position of the cart, their position inside the 3D space can change. However, the distance and the angle between the depth camera and the cart-mounted bins always stay the same, as they are firmly mounted on the cart. Considering HelmetPickAR, we were using a table that is divided into three areas for placing the picked items. As the target areas are on a table in the warehouse, the target areas stationary. Therefore, HelmetPickAR uses the same algorithm to detect if an object was placed in the target area than the one that is used for detecting the picks from bins. Overall, when a sphere was triggered, the system sends a trigger and the workflow is advanced accordingly or an error is displayed.



(a) Editor for trigger spheres (b) Arrows showing target compartments

Figure 7.3: The visual representation of the spheres that send a pick or place trigger (a). Hereby (1) are the stationary spheres from the shelves in the warehouse and (2) are the mobile spheres belonging to one side of the cart triggering when an item is placed for an order in a bin. (b) The arrows that indicating the direction of the next compartment to pick from.

7.4 Calibration

Before using both systems for the first time, three calibration steps have to be performed: setting up the model of the warehouse, calibrating the camera projector pairs, and calibrating the thresholds according to the user. For setting up the model of the warehouse, the trigger spheres for the environment have to be defined in the underlying 3D model of the warehouse to match the position of the compartments. In the OrderPickAR prototype, this is done using Unity3D. HelmetPickAR uses an XML-based approach, which stores the width, height, length, ID, and position of the compartments. Setting up the model of the warehouse has to be performed once when deploying the systems in a new environment. After this step, the calibration is stored and the systems can be used in the new environment. The trigger spheres have a unique ID, which are used to identify the picked items. In both systems, we are using a text file-based approach to load the workflow containing the orders to pick. However, this system can be easily integrated into an enterprise resource planning system, which could send and display orders in the moment they are issued by the customer. In the OrderPickAR system, the cart-facing cameras also have to be calibrated when adjusting the height of the cart to the warehouse. As the position of the camera changes when adjusting the height, the trigger spheres representing the bins in the cart (see Figure 7.3(a) (2)) have to be adjusted accordingly. After the initial calibration the distance between the cart-facing camera-projector pairs and the bins stay the same when moving the cart as the pair is firmly mounted on the cart's beam.

The second calibration task is calibrating the camera projector pairs. In both prototypes, we are using a simple four point calibration (as used by Hardy and Alexander [75]). Here, the projector displays four targets which are visible in the image seen by the camera. The user has to click on each target inside the camera's recorded image. Thereby, the system can store a mapping between the projector space and camera space. This procedure has to be done once for each of the camera-projector pairs.

In a last calibration step, the thresholds for triggering when a worker picks a part from a compartment or places a picked part at the target position have to be defined. This threshold is dependent on the length of the worker's arms, and the size of the worker's hands. E.g., workers with smaller hands will require a lower threshold for triggering a pick compared to workers with bigger hands. The length of the worker's arms is important for HelmetPickAR, as it needs a threshold for deciding if a change in the sensor data results from a worker's hand or is based on a change in the environment.

7.5 Displaying Visual Feedback

The location of the item to pick is usually not in the direct FoV of the worker. To reflect this fact in our setup, we deliberately chose the size of the warehouse to be larger than the field of view of the camera-projector pair facing the shelves on the cart or on the helmet. Therefore, we implemented feedback that is navigating the worker to the target compartment. In the OrderPickAR prototype the system projects an off-screen visualization (inspired by Baudisch and Rosenholtz [16]), which displays green arrows that are pointing towards the target compartment (see Figure 7.3(b)). The helmet prototype also uses arrows but navigates the user in a step-by-step approach from one compartment to the next. When the target compartment is inside the system's range of projection, the arrows disappear and the compartment is illuminated using a green light (see Figure 7.1(b) for the cart and Figure 7.2(b) for the helmet). We designed the visual feedback in a way that the whole compartment is illuminated using a simple color-based scheme. According to the color conventions that were defined in Section 5.3, we use green light to highlight the position of parts and red light to indicate when an error was made. To communicate the quantity of items to be picked, both systems also project the information directly into the compartment.

The visual in-situ feedback is calculated according to the currently loaded workflow. The workflow contains the ID of the target compartment's trigger sphere. Using the warehouse layout the ID reveals the position of the next target. Then arrows pointing towards the target can be calculated dynamically. When the target compartment is in the FoV of the systems, they use the projector to highlight the compartment using a green light. In case the user picks an item from a wrong compartment, the compartment is highlighted in red. After the user picked the order from the compartment, the systems differ in their visualizations. The OrderPickAR cart uses the projectors facing the cart to highlight the bin where the user should put the previously picked item using a green light. If the user puts an item into a wrong bin, the bin is highlighted in red. In contrast, HelmetPickAR projects arrows that show the worker in which direction the target bin is located. Similar to finding a compartment to pick from, once the target bin is in the FoV of the helmet, the bin is illuminated using a green light. It should be mentioned that both systems can only detect whether an item was picked from a compartment without knowing the quantity. Orders that require the worker to walk from the compartment to the target bin several times can be challenging. To support the worker when an order consists of too many items to carry from the source compartment to the target bin in one run, OrderPickAR implements a function that highlights the last order's compartment and target using a yellow light. Thereby, the user can find the shelves easier and finish the order using the yellow light although the feedback was already advanced. Afterwards the user can continue with processing the next order by following the green light. The yellow light is advanced when the following order is completed or when the cart is moved to another location. In the HelmetPickAR system, highlighting the previous work step using a yellow light is not implemented in the current version, however it could be easily integrated.

IV

Evaluation

Chapter 8

Evaluation of Feedback Mechanisms for Workplaces

This chapter is based on the following publications:

- M. Funk, A. Bächler, L. Bächler, O. Korn, C. Krieger, T. Heidenreich, and A. Schmidt. Comparing projected in-situ feedback at the manual assembly workplace with impaired workers. In *Proceedings of the 8th International Conference on PErvasive Technologies Related to Assistive Environments*, New York, NY, USA, 2015. ACM
- M. Funk, S. Mayer, and A. Schmidt. Using in-situ projection to support cognitively impaired workers at the workplace. In *Proceedings of the 17th international ACM SIGACCESS conference on Computers & accessibility*. ACM, 2015

When designing an assistive system for presenting in-situ instructions that are providing cognitive assistance during work processes there are many design alternatives. Instructions could show utility videos that are demonstrating how a work step should be performed, depict the steps in a pictorial instruction, or just highlight the important areas in the workplace. Especially for the user group

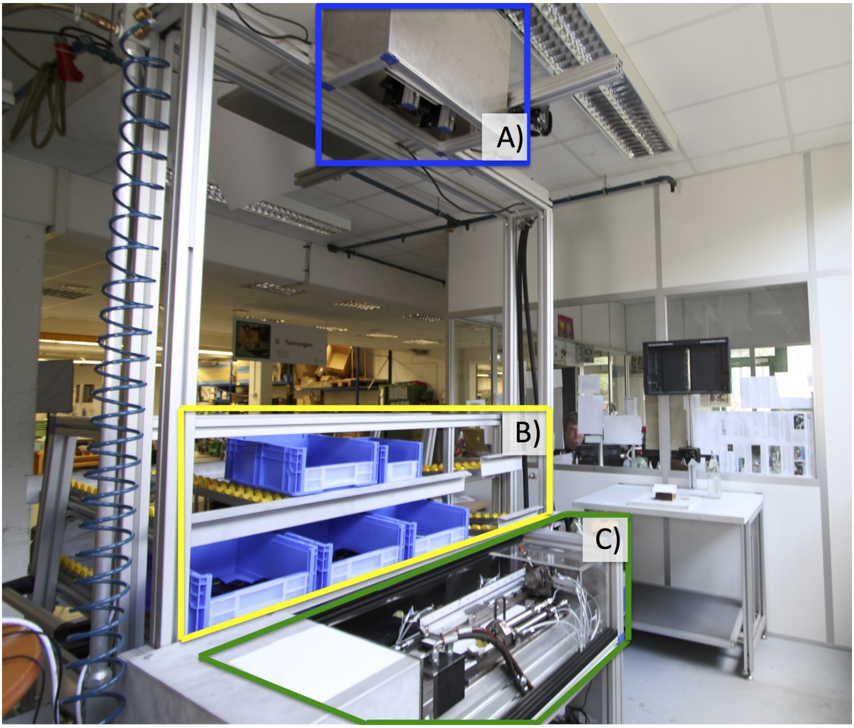


Figure 8.1: The assembly workplace is augmented with an assistive system consisting of a depth-camera and a projector (A). It enables displaying in-situ projected instructions during the work task directly onto the boxes holding the parts (B) and the work area (C).

of cognitively impaired workers, the way of presenting instructions is important to prevent workers from getting distracted during their tasks. In this chapter, we address the research questions RQ1 and RQ2. Accordingly, we first learn about what suitable in-situ visualizations for assembly instructions are (RQ1), and second, we learn about the benefit of in-situ assembly for cognitively impaired workers (RQ2). We address the research questions by performing two user studies in a controlled environment with cognitively impaired workers.

8.1 System and Visualizations

To compare different visualizations in a realistic setting, we installed our assistive system for a single workplace in an industry scenario which supports impaired workers during a real assembly task. Therefore, we equipped a machine that produces clamps with our camera-projector system. In the following, we introduce the prototype, the workflow of the clamp assembly, and the three visualizations.

8.1.1 Hardware Setup and Workflow

The assembly workplace (see Figure 8.1) is an automated machine that uses hydraulic and pneumatic components to assemble a clamp from five parts. The parts have to be inserted at a defined position in the machine. The machine is extended with the components of our assistive system. In this setup, which differed from our lab setup, the used components are a Microsoft Kinect_v1 depth camera and a Casio AJ-X251 projector. Both the projector and the depth camera are mounted on top of the machine to be able to project directly onto the work area and onto the picking bins that are holding the parts. Furthermore, the machine was designed in a way that each part can be placed from above and that both placed part and assembly position of the part are always visible for the camera to detect the work steps and for the projector to provide feedback.

The workflow for producing a clamp consists of picking and placing five parts, closing a safety glass, and opening the glass when the machine finished assembling the clamp. As picking and placing a part are considered as two different work steps, the whole workflow consists of twelve steps in total. We can differentiate between three types of work steps: picking a part out of a bin, placing a part in the machine, and special actions i.e. closing the safety glass.

8.1.2 Visualizations

We identified and extended three types of visualizations for projected instructions that can be understood by impaired workers from related work. Additionally, from interviews with supervisors of a sheltered work organization, we identified that new workflows are normally taught to impaired workers by a supervisor. Thereby, the supervisor lets the impaired worker perform the workflow and simultaneously gives verbal instructions. As the workers are used to getting additional verbal instructions, we provide pre-defined verbal instructions in addition to each projected feedback.

In spite of the fact that related work suggests using textual instructions for persons with mild cognitive disabilities [35], we could not conduct the experiment using textual instructions as many of the cognitively impaired workers participating in our study were not able to read.

Pictorial Instructions

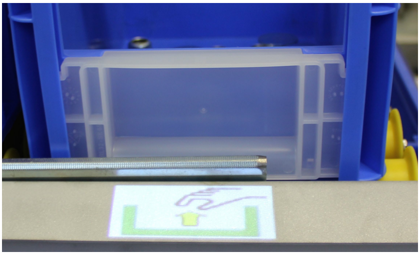
Work steps are often presented through pictorial instructions. For picking parts from the bins, we use a pictogram that is displayed directly in front of the bin from which the next part needs to be picked (see Figure 8.2(a)). The pictogram shows a hand that is picking a part from a bin and an arrow pointing away from the bin. For placing the part into the machine, we projected a full-sized picture of the part directly at the position where it needs to be placed (see Figure 8.2(d)). For closing and opening the safety glass, we projected a picture next to the glass, showing a hand that is moving the safety glass and an arrow that indicates the direction (see Figure 8.2(c)). As the whole machine did not fit in the worker's field of view when standing in front of it, the projected pictures were slightly blinking so that they could be spotted easily.

Video Instructions

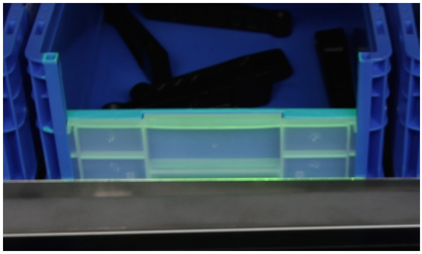
Another way to visualize instructions is projecting a utility video that is showing the work step that needs to be performed next. We chose to display short video clips that only show one action. We recorded the videos from the worker's point of view so that only little cognitive effort is needed to transfer the content of the video to the physical workplace (cf. utility videos in Section 2.5). For picking parts from the bins, a video showing the picking of the part is shown directly under the bin from where the worker picks parts. As the space for showing a video inside the work area is too small to accurately view the video, the system displayed the video of how to place a part at the white projection area next to the work area (see Figure 8.2(e)). The video showed where the part should be placed in the machine. For both opening and closing the safety glass, a different video is displayed at the projection area.

Contour Instructions

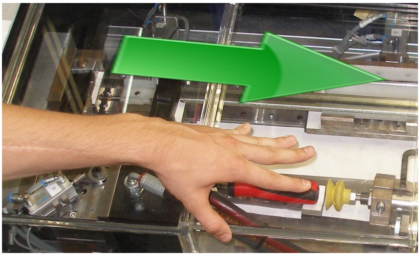
Previous work [3, 79] suggested that good instructions show the features of the parts that are changed in the current work step. Other features of the part that are



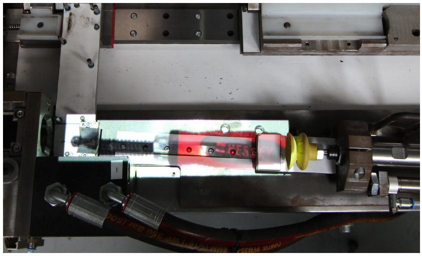
(a) Pictorial instruction for picking from bin



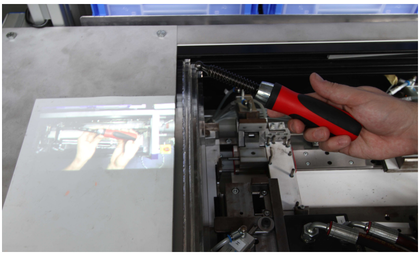
(b) Contour instruction for picking from bin



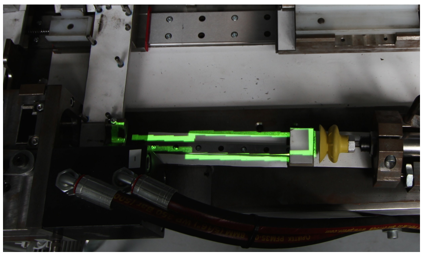
(c) Pictorial instruction used for closing the cover



(d) Pictorial instruction for placing a part



(e) Video instruction for placing a part



(f) Contour instruction for placing a part

Figure 8.2: An overview of the different visualizations that were used in the study. For picking a part from a box, (a) shows a pictorial instruction and (b) highlights the box as part of the contour visualization. For closing the safety glass, we projected the image shown in (c). The position of the part to place is visualized in (d) using a pictorial instruction, in (e) using a video, and in (f) using the contour visualization.

not relevant for the step should be simplified. Therefore, we decided to provide a pictorial instruction that just showed the contour of the current part. The contour conveys all relevant features, i.e. position and orientation, but hides potentially confusing details, e.g. printed text, color, and different material properties. In the contour visualization, for picking a part from a bin the system highlights the bin to pick from using a green light (see Figure 8.2(b)). When placing a part, the system displays the full-sized contour of the part at the correct position also using a green light (see Figure 8.2(f)). Closing and opening the safety glass is indicated by displaying a green arrow pointing to the direction the glass needs to be moved to. Just like in the pictorial visualization, we designed the contour visualization to be slightly blinking to enable the workers to spot the projection more easily.

8.2 Evaluation of Visualizations

As we are interested in evaluating the effect of the three introduced visualizations of in-situ projected instructions, we designed a between-subjects user study comparing proposed visualizations to a control group using no visual feedback. As participants, we invited cognitively impaired workers belonging to different Performance Index groups. As we assume that the visualization of the feedback is directly related to the perceived cognitive load and a regular NASA-TLX [77] questionnaire would be too complex for a cognitively impaired person to file, we use a simplified version of the NASA-TLX questionnaire. The simplified questionnaire is specially tailored to be understood by cognitively impaired participants (cf. [45] and Liane Bächler et al. (in prep.)). The questionnaire can be downloaded from the motionEAP project website[17]

INFO: This evaluation uses the simplified NASA-TLX questionnaire that was designed and proposed by Liane Bächler. The questionnaire is described in detail in the PhD thesis of Liane Bächler.

For evaluating the visualizations in a study with cognitively impaired workers, we designed the study following a between-subjects design with four groups using one of the three visualizations and a control group using no visual feedback. The only independent variable is the type of the projected instruction. As dependent

[17] The TLX for cognitively impaired participants is available on the motionEAP website: `http://www.motioneap.de/tlx-for-impaired-participants/`

variables, we measure the TCT, ER, and score of the simplified NASA-TLX for cognitively impaired participants.

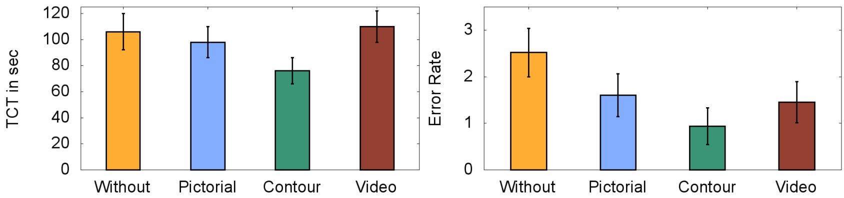
To create four groups which are equal in performance, we asked the supervisors of the cognitively impaired workers to assign each worker to one of three categories according to his or her Performance Index. The Performance Index is measured in percent and indicates to which extent an impaired worker is capable of performing a task considering assembly time and errors, compared to a non-impaired worker. The categories were defined by the supervisors of the sheltered work organization where we conducted the study. They were divided into groups as follows: Performance Index (PI) of 5%-10%, PI of 15%-35%, and a PI above 40%. Each group is assigned the same number of participants belonging to each PI group.

Participants

We recruited 64 participants (41 male, 23 female) for the study. The participants were aged from 16 to 59 years ($M = 41.7$, $SD = 10.6$). All participants were employees of a sheltered work organization and were either cognitively impaired or mentally impaired. None of the participants was familiar with the clamp-producing machine. Further, 67% of the participants had experience in manual manufacturing. The study took approximately 20 minutes for each participant.

Procedure

After welcoming the participant and explaining the course of the study, a general introduction about assistive systems for the workplace was given. At the beginning, each participant was assigned to one socio-educational instructor to support him or her during the task. This instructor stayed with the participant during the whole process to have a familiar person in this study scenario, which was new to the cognitively impaired participants. The instructor then helped the participants to complete an initial questionnaire collecting demographic information and prior experiences in manufacturing. Afterwards, the participant was assigned to one of the four conditions as described above. To familiarize the participant with the assembly of the clamp, the participant could assemble the clamp once using the visualization according to the assigned group while the instructor was giving additional verbal instructions. Afterwards, the participant had to assemble a clamp three times using only the visual feedback of the system according to the condition, or no feedback if the participant was assigned to the control condition. The system was recording the time that was needed for each assembly task automatically. The feedback was processed by a wizard of oz, who also counted the errors that were made during the assembly. After the



(a) Task completion time

(b) Error rate

Figure 8.3: (a) The task completion time according to the groups using different visualizations and (b) the error rate according to the visualizations used by the different groups. The error bars depict the standard error.

assembly was finished, the participant was guided to a calm area where they filled in the simplified NASA-TLX questionnaire. At the end additional qualitative feedback was collected. We randomly selected 13 participants to take part in a semi-structured interview, where we asked them about satisfaction, motivation, self-reliance and challenges caused by the usage of the system.

8.2.1 Results

Out of the 64 participants, two participants using the video condition were not able to complete the study because they were afraid of the projected videos. Thus, we excluded the two from the evaluation resulting in a total of 62 participants.

We statistically analyzed the TCT, ER, and the simplified NASA-TLX score between the groups using the different in-situ visualizations. The assumption of homogeneity of variance had not been violated ($p > .05$) for the TCT. A one-way ANOVA revealed no statistically significant effect on TCT between the groups ($F(3,58) = 1.446$, $p > .05$). The effect size estimate shows a medium effect ($\eta^2 = .069$). The group using the contour visualization was the fastest ($M = 76.52$s, $SD = 42.07$s), followed by the group using the pictorial visualization ($M = 98.13$s, $SD = 50.18$s) and the control group using no visual feedback ($M = 106.53$s, $SD = 56.74$s). The group using the video visualization took the longest time to assemble ($M = 110.45$s, $SD = 48.46$s). An overview of the TCT according to the different groups is depicted in Figure 8.3(a).

For analyzing the ER between the groups, we used a non-parametric Kruskall-Wallis-test as there were indications that the ER was not normally distributed.

Again, the assumption of homogeneity of variance had not been violated ($p > .05$). The Kruskall-Wallis-test did not reveal a significant difference $\chi^2(3) = 7.031$, $p = .071 > .05$. The effect size estimate shows a medium effect ($\eta^2 = .11$). The group using the contour visualization made the fewest errors ($M = .93$, $SD = 1.57$), followed by the group using the video visualization ($M = 1.45$, $SD = 1.75$), and the pictorial visualization ($M = 1.60$, $SD = 1.82$). The control group using no visual feedback made the most errors ($M = 2.52$, $SD = 2.07$). An overview of the ER according to the different groups is depicted in Figure 8.3(b).

We statistically compared the score of the simplified NASA-TLX between the groups using a non-parametric Kruskall-Wallis-ANOVA. As a post-hoc test, we used the Wilcoxon signed-rank test with an applied Bonferroni correction for all types of feedback, resulting in a significance level of $p < .0125$. The simplified NASA-TLX could be filled by all 62 impaired participants taking part in the study.

Mental demand. Considering the mental demand, the test revealed a significant difference $\chi^2(3) = 8.000$, $p = .046$. Pairwise Wilcoxon tests revealed that there is a significant difference between the group without visual feedback and the group using contour visualization ($Z = -2.572$, $p = .01$). The other pairwise tests did not reveal a significant difference (without vs. pictorial: $Z = -1.442$, $p = n.s.$, without vs. video: $Z = -.303$, $p = n.s.$, pictorial vs. contour: $Z = -1.033$, $p = n.s.$, pictorial vs. video: $Z = -1.117$, $p = n.s.$, contour vs. video: $Z = -2.197$, $p = n.s.$). The contour visualization was perceived the easiest ($M = .44$, $SD = .62$), followed by the pictorial visualization ($M = .81$, $SD = .98$), and the video visualization ($M = 1.21$, $SD = 1.05$). The group using no visual feedback perceived the task as most complex ($M = 1.31$, $SD = 1.01$) (see Figure 8.4 (A)).

Physical demand. Regarding the physical demand, the test could not reveal a significant difference between the groups using the different visualizations $\chi^2(3) = 4.278$, $p = n.s.$. The participants perceived the contour visualization as the least physically demanding feedback ($M = .44$, $SD = .62$), followed by the pictorial visualization ($M = .81$, $SD = .98$), and the video visualization ($M = 1.21$, $SD = 1.05$). The group using no visual feedback reported the highest physical demand ($M = 1.31$, $SD = 1.01$) (see Figure 8.4 (B)).

Temporal demand. The analysis of the perceived temporal demand did not reveal a significant difference between the groups using the different visualizations $\chi^2(3) = 3.161$, $p = n.s.$. As Figure 8.4 (C) shows, the contour visualization was perceived the fastest ($M = .50$, $SD = .73$), followed by the pictorial visualization ($M = .75$, $SD = .77$), and without in-situ feedback ($M = .87$, $SD = .71$). The

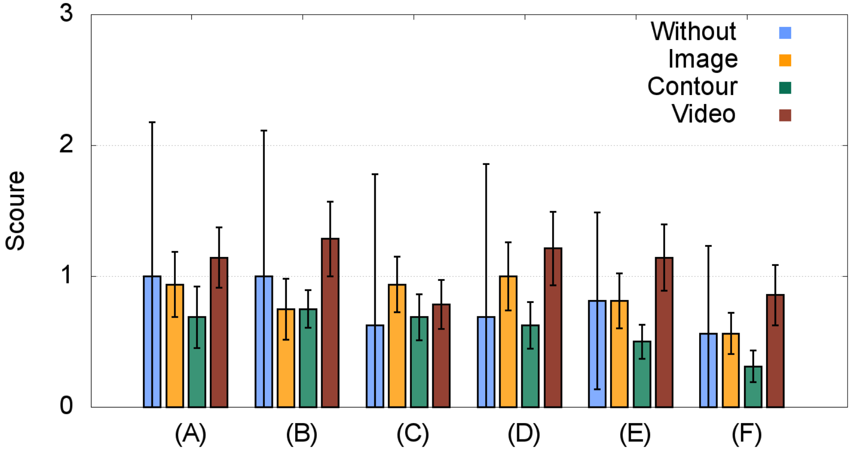
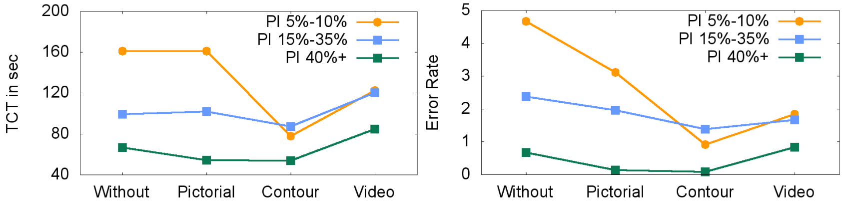


Figure 8.4: Overview about the results of the modified NASA-TLX. The error bars indicate standard error. (A) Mental demand, (B) Physical demand, (C) Temporal demand, (D) Performance, (E) Effort, (F) Frustration.

video visualization was perceived the most temporally demanding ($M = .93$, $SD = .82$).

Performance. We statistically compared the perceived performance of the participants between the different visualization groups. The test revealed a significant difference between the groups $\chi^2(3) = 8.493$, $p = .037$. Pairwise Wilcoxon tests revealed a significant difference between the group without visual feedback and the group using contour feedback ($Z = -2.575$, $p = .01$). The other pairwise tests did not reveal a significant difference (without vs. video: $Z = -.695$, $p = n.s.$, pictorial vs. video: $Z = -.794$, p=$n.s.$, contour vs. video: $Z = -2.433$, $p = n.s.$, pictorial vs. contour: $Z = -1.506$, $p = n.s.$, and without vs. pictorial: $Z = -1.287$, $p = n.s.$). Participants perceived their performance as best using the contour visualization ($M = .38$, $SD = .80$), followed by the pictorial visualization ($M = .81$, $SD = .98$), and the video visualization ($M = 1.00$, $SD = .78$). Participants using no visual feedback perceived their performance as least successful ($M = 1.31$, $SD = 1.13$). An overview is depicted in Figure 8.4 (D).

Effort. The statistical comparison of the participants' perceived effort between the different visualization groups did not reveal a significant difference $\chi^2(3) = 1.427$, $p = n.s.$. However, the participants perceived the least effort using the contour visualization ($M = .63$, $SD = .80$), followed by the pictorial visualization



(a) Task completion time (b) Error rate

Figure 8.5: (a) The task completion time according to the used visualization and the Performance Index (PI) of the participants. The contour visualization especially helps participants with the lowest PI. (b) The error rate according to the used visualization and the Performance Index (PI) groups. The contour visualization causes the fewest errors for all PI groups.

($M = .75$, $SD = .77$), and the video visualization ($M = .86$, $SD = .94$). As depicted in Figure 8.4 (E), the group using no visual feedback perceived their effort the highest ($M = 1.00$, $SD = .96$).

Frustration. Finally, we compare the frustration that was perceived by the participants using the different visualizations. The test revealed a significant difference between the groups $\chi^2(3) = 8.149$, $p = .043$. However, pairwise Wilcoxon tests did not reveal any significant differences between the different visualizations (without vs. video: $Z = -.183$, $p = n.s.$, pictorial vs. video: $Z = -1.569$, $p = n.s.$, contour vs. video: $Z = -2.388$, $p = n.s.$, without vs. contour: $Z = -2.334$, $p = n.s.$, pictorial vs. contour: $Z = -.821$, $p = n.s.$, without vs. pictorial: $Z = -1.467$, $p = n.s.$). Participants found the contour visualization the least frustrating ($M = .25$, $SD = .57$), followed by the pictorial visualization ($M = .44$, $SD = .72$), and the group using no visual feedback ($M = .75$, $SD = .68$). As depicted in Figure 8.4 (F), the group using the video feedback reported the highest frustration ($M = .86$, $SD = .86$).

We further analyzed the participants' performance according to the PI groups using a one-way ANOVA. Considering the ER, we found a significant difference ($F(2,59) = 7.251$, $p = 0.002$). A Tukey HSD post-hoc test revealed a significant difference between PI 40%+ and PI 5%-10%, as well as between PI 40%+ and PI 15%-35%. For the TCT, the test also found a significant difference ($F(2,59) = 7.999$, $p = 0.001$). Again, a Tukey HSD post-hoc test revealed a significant difference between PI 40%+ and PI 5%-10%, as well as between PI 40%+ and PI 15%-35%.

Finally, we compared the effect between the different visualizations and the PI groups regarding the TCT. A two-way ANOVA revealed a significant effect between the PI groups for the group using no visualization ($F(2,50) = 4.878$, $p = 0.012$) and the group using the pictorial feedback ($F(2,50) = 5.590$, $p = 0.006$). Pairwise comparisons revealed a significant difference between PI 5%-10% and PI 40%+, and between PI 5%-10% and PI 15%-35% for the group using no visual feedback. Regarding the pictorial feedback group, the test only revealed a significant difference between PI 5%-10% and PI 40%+.

Furthermore, we compared the effect between the different visualizations and the PI groups regarding the ER. A two-way ANOVA revealed a significant effect between the PI groups for the group using the pictorial visualization ($F(2,50) = 3.455$, $p = 0.039$) and the group using no visual feedback ($F(2,50) = 5.996$, $p = 0.005$). Pairwise comparisons reveal a significant difference between PI 5%-10% and PI 40%+ for the group using the pictorial visualization. Regarding the group using no visual feedback, the comparison reveal a significant difference between PI 5%-10% and PI 40%+, and between PI 5%-10% and PI 15%-35%.

The qualitative feedback revealed that the contour visualization was well perceived by the participants. P14 stated that he *"could see the alignment of the part from the shape of the projection"*. Regarding the video visualization, we observed that the frustration of the participants using the video was extremely high. We even needed to abort two experiments with the video visualization because the two participants were scared of the videos and started panicking. We could also observe that most participants using the video condition were not watching the video fully. They were just looking at it occasionally. P37 stated that he *"did not understand that the video tells me what to do"*. The participant rather inferred the position of the part based on the affordance of the part and the workplace.

In the semi-structured interview, we randomly selected 13 participants from different PI groups and different visualization groups. When we asked them about the general idea of visual feedback during the work process, 12 of the 13 interviewed participants stated that they felt adequately supported by the system. A total of 12 of the 13 participants experienced joy while working on the system. All participants feel an increased independence and self-responsibility in their work task, by using the system. E.g. P11 states that she *"is able to work confidently because of the flashing lights"*, and P5 states that he *"got help by the instructions and the lights of the system"*. Only 3 of the 13 interviewed participants feel challenged by working at the system. These participants expressed difficulties in learning and comprehending the instructions.

8.3 Discussion

The results of the study reveal that the contour visualization induces the fewest errors and the shortest TCT. Although the difference in committed errors and TCT is not statistically significant, we could observe a trend favoring the contour visualization. Regarding the TCT, the video visualization induced a longer TCT than the control group using no visual feedback. However, the difference is not statistically significant. A reason for this trend could be the additional time to watch the video.

Interestingly, the results show that the contour visualization induces significantly less perceived mental demand than the control group using no visual feedback. Furthermore, the perceived performance of the participants was significantly better using the contour visualization compared to the control group using no visual feedback. Overall, the contour visualization performed best in all six measures of the simplified NASA-TLX, however no significant differences could be found in the other four measures.

Considering the different PI groups, the results revealed a statistically significant difference regarding the TCT and the ER for the pictorial feedback and the control condition without visual feedback. For the video and the contour feedback, no statistically significant difference was found. This indicates that the contour feedback and the video feedback caused the PI groups to achieve results of similar quality (see Figure 8.5(a) and Figure 8.5(b)). This is particularly noticeable for the PI group 5%-10%, which achieves a TCT and an ER comparable to the other groups by using contour feedback. Thus, contour feedback might provide a means to enable a wide range of cognitively impaired workers to improve their performance.

Finally, visual support was generally well perceived by the participants. Many participants felt more confident when viewing projected instructions directly at the work place.

8.4 Towards Increasing Complexity

As the previous study showed, there is a benefit for all PI groups in using a simple contour-based visualization for providing in-situ instructions at the manual assembly workplace. We assume that this contour-based visualization is the best way to provide in-situ instructions for cognitively impaired workers using an assistive system. To analyze this effect further, we are interested in how well the contour-based in-situ instructions perform compared to state-of-the-art instructions and want to assess the capabilities of our assistive system. To make this comparison, we first want to learn about how many work steps a cognitively impaired worker is capable of performing using state-of-the-art instructions and later compare it to using our assistive system. For designing the experiment to reflect the current state-of-the-art, we first have to analyze how sheltered work organizations are usually organizing manufacturing in their factories and how instructions are presented. Traditionally, the assembly of products is split into small parts that the complexity of the assembly steps is very low and can be easily performed by cognitively impaired workers from each PI group. With a higher PI, the number of work steps that the cognitively impaired workers can perform in the sheltered work organizations increases. To analyze to which extent the assembly tasks are divided into sub tasks, we analyze one example factory of a sheltered work organization.

8.5 Analyzing the State-of-the-Art

To analyze the state-of-the-art of assembling products in a sheltered work organization, we analyzed a factory of the sheltered work organization GWW[18] with 72 impaired employees and 14 supervisors that are supporting the impaired employees. The supervisors consist of one social education supervisor and 13 technical supervisors. The impaired employees are workers with either cognitive disabilities (e.g. workers with down syndrome) or workers with mental disabilities (e.g. workers with tourette syndrome or burnout syndrome). The analyzed factory is producing cutting products (e.g. scissors, pliers, or pincers) and further has a carpentry for producing tables and benches.

The studied factory offers 93 manual assembly workplaces. We counted the number of assembly steps that are preformed at each workplace. This count includes different types of work steps, e.g. picking up and placing of a part, or using a tool. According to this method, the analyzed workplaces consist of 1 to 25 work steps per workplace. Our analysis revealed an average amount of 5.25 ($SD = 4.05$) work steps per workplace.

To support workers during their tasks, the sheltered work organization offers pictorial instructions that are mounted directly over the bins which hold the parts

[18] GWW - Gemeinnützige Werkstätten und Wohnstätten GmbH - `http://www.gww-netz.de/` - last access 5th Oct. 2016

Figure 8.6: A state-of-the-art assembly workplace that is used in the analyzed sheltered work organization. Pictorial instructions above the bins holding the parts to assemble show how the assembly has to be performed. Workers can compare their assembled product with the depicted instruction.

to be assembled (see Figure 8.6). The instructions show the assembled product in their intermediate state after the part in the bin is assembled. The workers can control their assembled product using the picture, e.g. if the last part was assembled correctly or where to assemble the next part. In case the workers do not understand the pictorial instructions, there is always a supervisor around, who can provide help with assembling the next part.

Overall, the factory is designed to split each product into small sub tasks, which are easy enough so that the impaired workers can perform them just with the help of the pictorial instructions. This segmentation of work tasks leads to a higher level of satisfaction of the workers because they are able to complete the whole task without help from a supervisor. Just in case they need help, they are able to consult the supervisors.

8.6 Evaluation of Task-Complexity

For comparing the state-of-the-art pictorial instructions to in-situ projected instructions, we conducted a repeated measures user study using our assistive system for a single workplace that was introduced in Chapter 6. Informed by previous work [138, 149], we chose a Lego Duplo task as an abstract assembly task that can be easily scaled up to provide more work steps without introducing a different product. Furthermore, such a pick-and-place task is a good abstraction of tasks that are usually performed in sheltered work organizations, as those tasks also require picking parts and placing them at defined assembly positions. However, using a tool on the placed assembly parts is not included in this abstract assembly task.

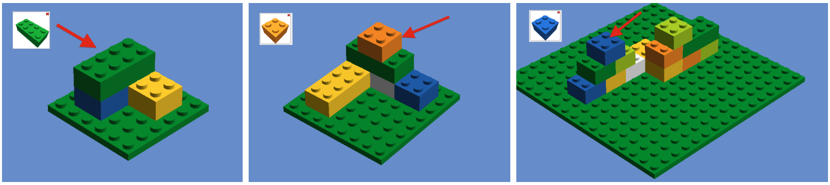
8.6.1 Method

For evaluating the system, we used a repeated measures design with two independent variables: The used instruction, and the number of bricks in the assembled construction. We measure the TCT and the ER as dependent variables. To normalize the data, we divide the TCT and the ER by the number of bricks of which the construction consists to get the time per brick (TPB) and the errors per brick (EPB).

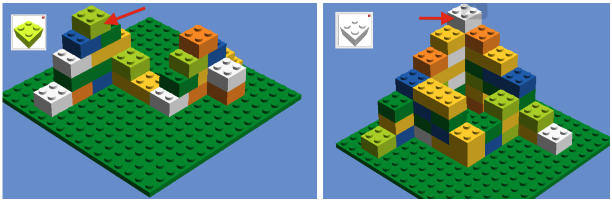
Apparatus

We consider 5 different difficulty levels with constructions consisting of different numbers of bricks (see Figure 8.7): 3, 6, 12, 24, and 48 bricks. As placing one brick results in two work steps, i.e. picking the brick and placing it at the correct position, the levels result in 6, 12, 24, 48, and 96 work steps. The Lego Duplo constructions in their correctly assembled state are depicted in Figure 8.7 (a)-(e).

For the in-situ instruction condition, we used our single workplace setup of our assistive system described in Chapter 6, which highlights the box to pick from and displays the contour of the picked brick at the correct assembly position. As the assembly detection requires the participant to remove his or her hands from the assembled brick, the researcher is able to advance the feedback manually using a wireless presenter in case the participant is occluding the assembled part by leaving the hands above the assembled brick.



(a) 3 brick construction (b) 6 brick construction (c) 12 brick construction



(d) 24 brick construction (e) 48 brick construction

Figure 8.7: The constructions used in the study. We consider five different complexity levels: (a) 3 bricks, (b) 6 bricks, (c) 12 bricks, (d) 24 bricks, and (e) 48 bricks. The images depict the final step of the pictorial instructions.

As a state-of-the-art control condition, we use pictorial instructions to show the next part to pick and the assembly position to the worker. We use a 28" screen that is placed next to the assembly area (see Figure 6.1). The pictorial instructions provide three main types of information. First, the type of brick to pick, depicted by a icon in the upper left corner (also see Figures 8.7 (a)-(e)). Second, a picture showing the work-piece in the correctly assembled state after the current brick was assembled at the correct position. Third, a red arrow directly highlighting the position of the last placed brick. This type of information enables the participant to see the placement position of the current brick immediately. This is useful because finding the correct position of the brick could be cumbersome, especially with an increasing number of bricks. The instructions were created using the Lego Digital Designer[19]. The pictorial instructions were proceeded by the researcher using a wireless presenter after the participant placed the brick at a position. This enabled the participant to fully focus on the assembly task.

[19] Lego Digital Designer - `http://ldd.lego.com/en-us` - last access 5th Oct. 2016

Procedure

After explaining the purpose of the study, we asked the participant to sit in front of the assistive system. Depending on the condition to be conducted, we either explained how a pictorial instruction is understood or how the in-situ projection shows the part to pick and where to assemble it. We instructed the participants to primarily focus on assembling the constructions correctly and only secondarily focus on the assembly time. During the experiment, the TCT was taken and the number of errors was counted independently by two researchers. The researchers started the measuring of the TCT upon showing the first instruction and stopped the measuring when the construction was finished. In case of inconsistency between the two counted error numbers, the assembled building was analyzed for errors. The researchers counted picking errors and placement errors. A picking error is counted when the participant was picking a brick from a wrong box, and a placement error is counted when a brick is placed at a wrong position in the assembly area considering the relative position to the other bricks. In case a placement error effected the possibility to finish the construction correctly and not to influence further work steps, the researchers paused the experiment and the measuring of the TCT to get the assembly back into a correct state. In the pictorial instruction condition, the absolute position on the plate was not checked for correctness. The researchers instructed the participants to begin with the first brick in the middle of the plate and not determine the exact position of the brick on the plate in the pictorial instruction. This procedure was repeated for all five constructions for both the pictorial and in-situ conditions respectively. After each condition, subjective feedback from the participant was collected by asking for their opinion about the feedback of the respective conditions. The order of the constructions and the conditions were counterbalanced according to the Balanced Latin Square over the 15 participants. We ensured that each Performance Index group had the same five orders of the constructions.

Participants

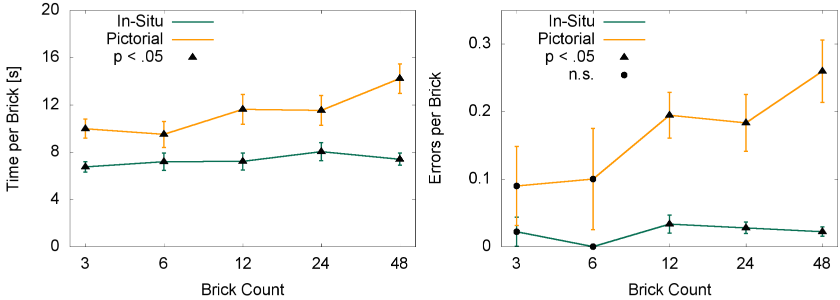
As previously described, the sheltered work organization where we conducted the study uses a Performance Index to assess the performance of their workers and to be able to assign them to tasks that the workers are capable of conducting. The PI is measured in percent and indicates to what extent the impaired worker is capable of performing a task compared to a worker without disabilities. The PI is determined subjectively by the supervisor of the impaired worker who works with the impaired worker every day. We considered three PI groups: PI of $5\% - 10\%$, PI of $15\% - 35\%$, and a PI over 40%. We chose the participants for the study in a way that five participants belonging to each PI group took part in the study, which results in a total number of 15 participants.

Accordingly, we recruited 15 participants (4 female, 11 male) for the study. The participants were aged from 20 to 55 years ($M = 40.1$, $SD = 10.33$). All participants were employees of a sheltered work organization and were workers with a cognitive disability. None of the participants were familiar with the Duplo constructions that were assembled in the study. However, all participants had experiences playing with Duplo bricks before. For each participant, the study took approximately 60 minutes.

8.6.2 Results

We statistically analyzed the TPB and the EPB between the in-situ instructions and the pictorial instructions using a one-way repeated measures ANOVA. The assumption of homogeneity of variance had not been violated ($p > .05$) for the TPB and the EPB.

Considering the TPB for the task consisting of 3 bricks, the in-situ instructions were faster ($M = 6.74$s, $SD = 1.72$s) than the pictorial instructions ($M = 9.98$s, $SD = 3.26$s). The analysis revealed a significant difference between the instructions ($F(1,14) = 18.088$, $p < .001$). The effect size estimate shows a large effect ($\eta^2 = .564$). For the task consisting of 6 bricks, the in-situ instructions were faster ($M = 7.18$s, $SD = 2.95$s) than the pictorial instructions ($M = 9.49$s, $SD = 4.41$s). The ANOVA revealed a significant difference between the instructions ($F(1,14) = 5.698$, $p = .032$). The effect size estimate shows a large effect ($\eta^2 = .289$). When analyzing the 12-brick task, again the in-situ instructions were faster ($M = 7.20$s, $SD = 2.93$s) than the pictorial instructions ($M = 11.61$s, $SD = 5.01$s). The statistical comparison revealed a significant difference between the instructions ($F(1,14) = 22.567$, $p < .001$). The effect size estimate shows a large effect ($\eta^2 = .617$). For the task consisting of 24 bricks, the in-situ instructions were faster ($M = 8.03$s, $SD = 3.09$s) than the pictorial instructions ($M = 11.53$s, $SD = 5.05$s). The ANOVA revealed a significant difference between the instructions ($F(1,14) = 12.981$, $p = .003$). The effect size estimate shows a large effect ($\eta^2 = .481$). Finally, when analyzing the task consisting of 48 bricks, the in-situ projected instructions were faster ($M = 7.40$s, $SD = 2.00$s) than the pictorial instructions ($M = 14.21$s, $SD = 5.04$s). The ANOVA revealed a significant difference between the instructions ($F(1,14) = 50.027$, $p < .001$). The effect size estimate shows a large effect ($\eta^2 = .781$). Figure 8.8(a) shows an overview of the results.



(a) Time per brick (b) Errors per brick

Figure 8.8: (a) Overview showing the time needed to pick and assemble a brick dependent on the complexity of the product to assemble. (b) Overview showing the errors made dependent on the complexity of the product to assemble. Error bars depict the standard error. A triangle indicates a significant difference between the conditions.

Considering the EPB for the task consisting of 3 bricks, the in-situ instructions led to less errors ($M = 0.02$, $SD = 0.08$) than the pictorial instructions ($M = .08$, $SD = .23$). The analysis did not reveal a significant difference between the instructions ($F(1,14) = 1.014$, $p = n.s.$). For the task consisting of 6 bricks, the in-situ instructions did not lead to any error. The pictorial instructions lead to a few errors ($M = 0.10$, $SD = 0.30$). The ANOVA did not reveal a significant difference between the instructions ($F(1,14) = 1.669$, $p = n.s.$). When statistically analyzing the EPB of the 12-brick task, the in-situ instructions ($M = 0.03$, $SD = 0.05$) lead to less errors than the pictorial instructions ($M = 0.19$, $SD = 0.13$). The statistical comparison revealed a significant difference between the instructions ($F(1,14) = 27.605$, $p < .001$). The effect size estimate shows a large effect ($\eta^2 = .664$). For the task consisting of 24 bricks, the EPB of the in-situ instructions were lower ($M = .02$, $SD = .03$) than the EPB of the pictorial instructions ($M = 0.18$ sec, $SD = .16$). The ANOVA revealed a significant difference between the instructions ($F(1,14) = 15.321$, $p = .002$). The effect size estimate shows a large effect ($\eta^2 = .523$). Finally, when analyzing the task consisting of 48 bricks, the projected instructions lead to less errors ($M = .02$, $SD = .02$) than the pictorial instructions ($M = .26$ sec, $SD = .18$ sec). The ANOVA revealed a significant difference between the instructions ($F(1,14) = 30.455$, $p < .001$). The effect size estimate shows a large effect ($\eta^2 = .685$). Figure 8.8(b) shows a graphical representation of the results.

We further analyzed the effect of number of work steps on TPB and EPB for both pictorial and in-situ instructions using a repeated measures ANOVA. Mauchly's test showed that the sphericity assumption was not violated for TPB and EPB.

Regarding the pictorial instructions, the analysis revealed a significant difference in TPB between the different step sizes ($F(4,56) = 7.144$, $p < .001$). Pairwise comparisons revealed that the difference in TPB between the 48 brick and 12 brick task, 48 brick and 6 brick task, and 48 brick and 3 brick task are significantly different (all $p < .05$). Considering the EPB, between the different step sizes using pictorial instructions, the analysis did not reveal a significant difference ($F(4,56) = 2.291$, $p = n.s.$). Considering the in-situ instructions, the analysis did not reveal a significant difference in TPB between the different step sizes ($F(4,56) = 1.264$, $p = n.s.$). Also for the EPB, the ANOVA could not show a significant difference between the step sizes ($F(4,56) = 1.012$, $p = n.s.$).

During the study, the participants commented on the different types of instructions. Regarding the in-situ instructions, participants liked that *"the system is showing the next box"* (P7, P11) and that *"[it] exactly shows where to put the next brick"* (P3, P7, P14). A participant referred to the system as *"magic light that helps performing the task"* (P2). Considering the pictorial instructions, the participants liked that *"the instructions are shown on a computer rather than on a printout"* (P7). However, a participant stated that he was *"having problems to find the correct positions as other bricks in the image are confusing"* (P6).

The supervisors of the sheltered work organization reported that in the days after the study, the participants were asking them if they can work at the *"workplace with the lights again"* and that it was fun for them and they enjoyed working with our system. They even asked when they will be able to perform their regular tasks with the help of *"the lights."*

8.7 Discussion

The results of the user study suggest that in-situ instructions have several advantages over the state-of-the-art pictorial instructions. First, the time per brick is up to 1.6 times lower using the in-situ instructions. The difference between the in-situ and pictorial instructions is statistically significant for all used complexity levels that were used in the study. Second, the errors per brick is up to 3 times lower using the in-situ instructions compared to pictorial instructions. This difference is statistically significant for the constructions consisting of 12, 24, and 48 bricks. When considering the TPB and EPB, the values across the different complexity levels are linear for the in-situ instructions. For the pictorial instructions, the difference between the complexity levels regarding the TPB is significantly different. The qualitative feedback also indicates that the participants preferred the in-situ instructions over the pictorial instructions as the in-situ instructions were always showing the position. Considering the pictorial instructions, the participants found that with increasing complexity it becomes harder to find the correct assembly position. This was despite the red arrow indicating the position. We assume that this is because a more complex structure requires more cognitive processing and that the assembly position is only highlighted in the instruction and not in the assembly position in the physical workplace.

8.7.1 Implications

The aforementioned user study revealed two implications considering the design of assembly tasks for cognitively impaired workers in sheltered work organizations. First, in-situ projected instructions should be used to instruct workers rather than pictorial instructions. When using in-situ instructions instead of pictorial instructions, impaired workers could assemble faster and with fewer errors. Second, impaired workers could assemble more complex products with a steady error rate and a steady assembly time, even with increasing complexity of the work task. This could further integrate impaired workers into the working life and could lead to a higher satisfaction of the cognitively impaired workers.

8.7.2 Limitations

It should be mentioned that the proposed system has certain limitations. Some of the impaired workers in the user study were leaving their hand in the work area covering the previously assembled brick which caused the system not to trigger automatically. Therefore, we were using a wireless presenter to advance the feedback manually in case the workers occluded the assembled part and retained covering the bricks. We also discovered that the Kinect_v2 sensor needs to run warm first before it can accurately detect correct assembly. When teaching the reference values with a Kinect that was recently started, the data became invalid after 20 minutes. Our observations suggest starting the Kinect 45 minutes before using it for assembly detection.

8.7.3 Public Exhibition

To show our assistive system using in-situ projection to a broader audience, we exhibited our system at the trade fair for vocational rehabilitation and exhibition of workshops for persons with disabilities[20] in Nürnberg, Germany. During the four days of the fair, over 400 impaired persons were able to try our assistance system. We mounted our prototype on a height-adjustable table, to enable persons using a wheelchair to use our system, too. As a demo scenario, we considered assembling a Lego Duplo wall consisting of nine different bricks resulting in 18 work steps. We received positive feedback throughout the demo from both impaired persons trying our system as well as supervisors working for sheltered work organizations. Visitors stated that the system was *"easy to learn and use"*. Another visitor liked that he *"just needs to focus on one thing to have a positive achievement."*

8.8 Summary

In this chapter, we evaluated which visualization for in-situ assembly instructions is most appropriate to provide cognitive support for workers with cognitive impairments. Further, we analyzed to which extent in-situ instructions outperform traditional pictorial step-by-step instructions.

Considering the visualizations of in-situ instructions, our results favor a simple contour-based visualization, which can easily be understood by cognitively impaired workers. Through a user study involving 64 cognitively impaired participants, we compared a contour visualization, a pictorial visualization, and a video visualization in a real world work scenario, i.e. assembling a clamp. The results of the study reveal, that the perceived performance is significantly higher and the perceived mental load is significantly lower using the contour visualization compared to using no visual feedback. Further, the results indicate that the contour visualization performs best concerning error rate and task completion time, however not statistically significant. Another interesting observation made in the study, is that video visualizations of instructions might have a negative effect on cognitively impaired workers, as the task completion time was higher compared to the other visualizations, however also not statistically significant.

[20] `https://www.werkstaettenmesse.de` (last access 5th Oct. 2016)

Through a second user study with 15 cognitively impaired workers, we compared different levels of complexity in assembled products using traditional pictorial step-by-step instructions and contour-based in-situ instructions. The results of the study show that in-situ instructions lead to faster assembly times and lead to less errors compared to state-of-the-art pictorial instructions. This effect is statistically significant for tasks consisting of 3, 6, 12, 24 and 48 assembled parts when considering the assembly time and for tasks consisting of 12, 24, and 48 assembled parts when considering the number of errors. These results might have a great impact on how tasks are divided into workplaces at sheltered work organizations. Especially as using an assistive system with in-situ projection could empower cognitively impaired workers to work on more complex tasks and thereby foster inclusion.

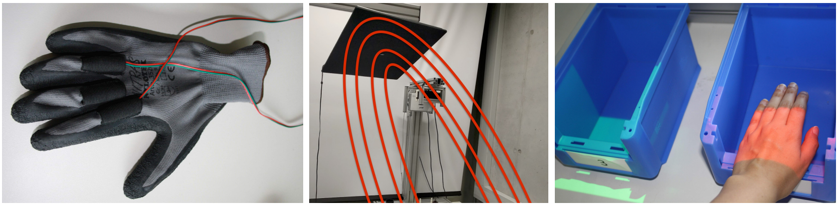
Chapter 9

Evaluation of Error Feedback for Workplaces

This chapter is based on the following publication:

- M. Funk, J. Heusler, E. Akcay, K. Weiland, and A. Schmidt. Haptic, auditory, or visual? towards optimal error feedback at manual assembly workplaces. In *Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, New York, NY, USA, 2016. ACM

Our assistive system for the workplace and the assistive systems described in related work mostly use visual feedback for providing assembly instructions or notifying the worker in case an error was made. However, a red light indicating an error might not always be the best solution for communicating that an error was made, or might be overlooked in stressful situations. Therefore, we extended our assistive system to compare haptic, auditory, and visual error feedback at the manual assembly workplace and tried to find the best way of communicating an error using an assistive system (RQ3). Through a pre-study with nine participants and a main study with 16 participants, we first determine suitable variants for each error feedback modality, and second, compare the error feedback modalities against



(a) Haptic feedback (b) Auditory feedback (c) Visual feedback

Figure 9.1: (a) To provide haptic error feedback, we mounted vibration motors on a regular safety glove. (b) For providing auditory error feedback, we use a Holosonics AS24i directed sound speaker. The red lines indicate the area where the error sound is noticeable. (c) For the visual error feedback, we use our assistive system highlighting the area where an error was made using a red light.

each other. The results show that haptic feedback is appropriate for retaining the worker's privacy, and auditory feedback is perceived as most distracting. The subjective feedback reveals interesting insights for future research opportunities, as participants rated a combination of visual and haptic feedback as appropriate.

9.1 Apparatus

For comparing the haptic, auditory, and visual error feedback, we extended the laboratory test setup (described in Chapter 6) of our assistive system for a single workplace to provide the three modalities of error feedback (see Figure 9.1). To achieve that, we integrated a glove providing haptic error feedback and a speaker providing auditory error feedback as external components, which listen to the error trigger (cf. Chapter 5).

9.1.1 Haptic Feedback

Inspired by previous work [20], for providing haptic feedback when an error was made, we equipped a regular work safety glove with two vibration motors (see Figure 9.1(a)). The motors are positioned at the upper end of the index finger and the upper end of the ring finger. The motors are powered and controlled using a wire that is mounted at the worker's upper forearm to not interfere with the assembly task. The wires are connected to an Arduino Yun micro controller which listens to the error trigger of the assistive system via UDP. The trigger is only sent when the worker picked an item from the wrong bin as assembly errors were not detected in this experiment. The haptic feedback provided by the gloves has the benefit that error feedback can only be perceived by the worker wearing the glove. This retains the worker's full privacy.

9.1.2 Auditory Feedback

When designing auditory feedback for the workplace, two factors have to be considered: companies' safety regulations and workers' privacy. Auditory error feedback that retains users' privacy is, e.g., playing auditory error feedback in headphones that the user is wearing. However, due to safety regulations of most companies, wearing headphones is not allowed while performing assembly tasks. The usage of regular speakers would be allowed by the companies. However, persons that are standing near the worker would also hear that the worker made a mistake. Therefore, we decided to use the Holosonics Audio Spotlight 24i[21] (AS24i), which provides directed sound that is only noticeable at the workplace where the error occurred and which does not require the worker to wear any piece of technology. As the sound is kept at the workplace using ultrasonic waves, the sound is only perceived according to the sketch in Figure 9.1(b).

9.1.3 Visual Feedback

For the visual feedback, we are using the in-situ projection of our assistive system, which is highlighting the assembly position and the bins to pick from. Considering the colors, we use the conventions discussed in Section 5.3. Accordingly, green light is used to communicate the position of the next picking bin and where to assemble the picked part. Red light is used to communicate that the user tried to pick from a wrong picking bin. Figure 9.1(c) shows the red error feedback and the green instruction highlighting the box to pick from.

---

[21] `http://www.holosonics.com/products.html` - last access 5th Oct. 2016

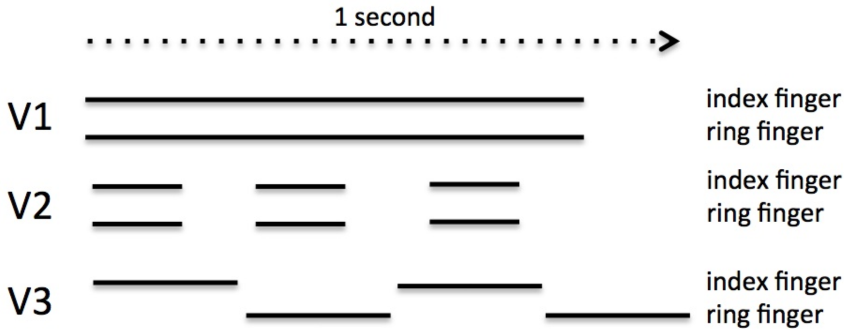


Figure 9.2: A graphical representation of the three variants of the haptic feedback that were used in the pre-study.

9.2 Pre Study: Determining Suitable Variants

As the perception of the error feedback modality is also dependent on designing suitable variants of error feedback for each modality, we designed three variants for each error feedback modality inspired by related work and preliminary tests.

For the haptic feedback using the vibration motors, we considered three variants that are depicted in Figure 9.2. The first variant (V1) is a constant vibration on both fingers, the second (V2) is three short vibrations on both fingers, and the third variant (V3) is alternately vibrating on both fingers. Considering the auditory feedback, we were using a deep error tone, a high error tone, and a major-triad sound. For the visual feedback, we were using a static red light, a blinking red light, and a pulsing red light. The blinking was designed in a way that the light was 500ms visible and 300ms off. The pulsing took 500ms to increase the intensity of the light and 500ms to decrease the intensity.

9.2.1 Method

To find the most suitable variant for each modality, we conducted a pre study following a repeated measures design with the used variant as the only independent variable. As dependent variable, we asked the participants to rate the variant on a 7-point Likert scale considering appropriateness for error feedback. We counterbalanced the order of the variants and the order of the modalities.

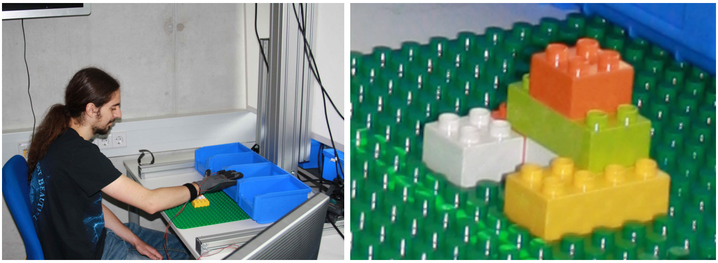
After explaining the course of the study and collecting the demographics, we asked the participants to sit at the assembly workplace. For each error feedback variant, we asked the participants to pick a part from a wrong bin three times. Afterwards, we asked the participant to rate the feedback on a 7-point Likert scale. We repeated the procedure for the other variants and modalities.

For the study, we invited 9 participants (1 female) aged between 19 and 28 years ($M = 22.3$, $SD = 2.58$). The participants were students with various majors, who were recruited via our mailing list. We compensated them for the participation with candies. The study took approximately 30 minutes per participant.

9.2.2 Results

Considering the haptic feedback variants, participants considered V3 the best ($M = 5.5$, $SD = .95$), followed by V2 ($M = 4.6$, $SD = 1.63$) and V1 ($M = 3.5$, $SD = 1.4$). A one-way repeated measures ANOVA could not reveal a significant difference between the variants ($p > .05$). For the auditory feedback variants, participants gave the best rating to the deep error tone ($M = 4.6$, $SD = 1.49$), followed by the high tone ($M = 3.53$, $SD = 2.06$), and the triad ($M = 3.44$, $SD = 1.25$). A one-way repeated measures ANOVA could not reveal a significant difference between the variants ($p > .05$). Lastly, for the visual feedback variants participants rated the static red light ($M = 5$, $SD = 1.24$) and the blinking red light ($M = 5$, $SD = 1.34$) with the same score. The pulsing red light was rated the worst ($M = 3.11$, $SD = .99$). A one-way repeated measures ANOVA revealed a significant difference, $F(2,16) = 10.321$, $p < .001$. Post hoc tests revealed that the pulsing red light was perceived significantly worse than the other two variants (all $p < .05$).

Based on the results, we considered using the alternately vibrating pattern (V3) as haptic error feedback, the deep error tone as auditory error feedback, and the static red light as visual error feedback.



(a) Participant using the study setup (b) Lego Duplo assembly task

Figure 9.3: (a) The workplace used in the study consists of an assembly plate and 4 picking bins with Lego Duplo bricks. A participant is using the glove providing haptic error feedback. (b) We used a 6 brick Lego Duplo construction as assembly task in the user study.

9.3 Main Study: Towards the Optimal Error Modality

After empirically determining suitable variants for each modality, we conducted a user study which compares the three modalities (haptic, auditory, and visual) as well as combinations of the three modalities for providing error feedback at a manual assembly workplace.

9.3.1 Method

We conducted the study following a two-steps repeated measures design with the used feedback modality as the only independent variable. For the first step, we only compared the haptic, auditory, and visual feedback. We counterbalanced the order of the modalities for each participant. As dependent variables we measured the user's subjective rating using a 7-point Likert scale considering overall rating, privacy, and distraction. As second step, we conducted the study with combinations of the three modalities: haptic and auditory, haptic and visual, auditory and visual, and all modalities simultaneously. As dependent variable we collected the overall rating using a 7-point Likert scale. Further, we collect qualitative feedback through semi-structured interviews.

After explaining the course of the study and signing a consent form, we collected the demographics and gave the participants a general introduction about the assistive system and the task. We instructed the participants to take a seat at the laboratory test setup of our single assembly workplace that is depicted in Figure 9.3(a). Inspired by related work [149], we considered assembling a 6 brick Lego Duplo construction (see Figure 9.3(b)). Throughout the study, the system only displayed the assembly position using a green rectangle. We instructed the participants to place the picked brick directly at the position that was highlighted by the system. For picking bricks from the four picking bins, the system did not provide any instructions. As we wanted the participants to make mistakes and to pick items from a wrong bin, we instructed them to find the correct bin by trying. If the participant picks from a wrong bin, the error feedback according to the condition is triggered. If the correct bin is picked from, the assembly position is highlighted by the system. After assembling the 6 bricks of the construction, we asked the participant to fill a questionnaire containing the 7-point Likert scales. After completing the three modalities, we tested combinations of the modalities. Finally, we conducted a semi-structured interview for collecting qualitative feedback. The study approximately took 45 minutes per participant.

9.3.2 Participants

For the study, we recruited 16 participants (all male) aged from 19 to 24 (M = 20.53, SD = 1.15). The participants were recruited via mailing list and were students with various majors. None of the participants was familiar with assistive systems for the workplace. They used the error feedback for the first time. The participants received candies for participating in our experiment.

9.3.3 Results

Throughout the study, each participant made at least 8 errors on average using each modality or combination of modalities. We statistically compared the responses of the 7-point Likert scales using a repeated measures ANOVA. Mauchly's test showed that the sphericity assumption was violated for privacy ($\chi^2(2)$= 6.29, p = .043). Therefore, we used the Greenhouse-Geisser correction to adjust the degrees of freedom (ε=.734). We used a Bonferroni correction for all post-hoc tests.

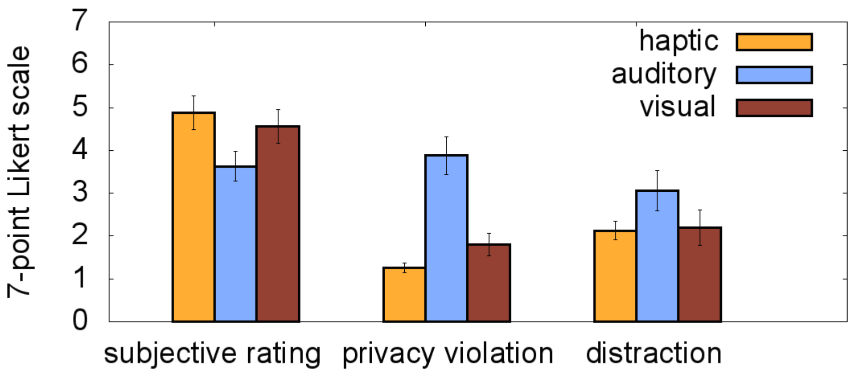


Figure 9.4: The results of how the three feedback modalities were perceived by the participants in the user study. The error bars depict the standard error.

First, we analyzed the average overall rating of the feedback modalities (1 = very bad and 7 = very good). The haptic feedback was perceived best ($M = 4.88$, $SD = 1.58$), followed by the visual feedback ($M = 4.56$, $SD = 1.59$), and the auditory feedback ($M = 3.63$, $SD = 1.36$). A one-way repeated measures ANOVA did not reveal a significant difference considering the overall rating of the modalities ($p > .05$). However, the effect size estimate shows a large effect ($\eta^2 = .161$). A graphical representation is depicted in Figure 9.4.

Second, analyzing to which extent the participants felt that their privacy was violated (1 = not at all and 7 = very much), the haptic feedback was rated best ($M = 1.25$, $SD = .45$), followed by the visual feedback ($M = 1.81$, $SD = 1.05$), and the auditory feedback ($M = 3.88$, $SD = 1.75$). A one-way repeated measures ANOVA revealed a significant difference between the conditions, $F(1.469, 22.028) = 24.864$, $p < .001$. The post-hoc test showed a significant difference (all $p < .05$) between the haptic and the auditory feedback, and between the visual and the auditory feedback. The effect size estimate shows a large effect ($\eta^2 = .624$). Figure 9.4 depicts the results graphically.

Considering how distracted the participants felt by the different feedback modalities (1 = not at all and 7 = very much), the haptic feedback was rated best ($M = 2.13$, $SD = .89$), followed by the visual feedback ($M = 2.19$, $SD = 1.64$), and the auditory feedback ($M = 3.06$, $SD = 1.91$). However, a one-way repeated measures ANOVA did not reveal a significant difference for the perceived distraction ($p > .05$). The effect size estimate shows a medium effect ($\eta^2 = .098$). Figure 9.4 shows a graphical representation of the results.

Finally, we analyze the participants' overall rating of the combinations of the modalities (1 = very bad and 7 = very good). The combination that was rated the best is the haptic and visual feedback ($M = 4.50$, $SD = 1.46$), followed by the auditory and visual feedback ($M = 3.88$, $SD = 1.2$), and the auditory and haptic feedback ($M = 3.56$, $SD = 1.71$). Combining all three modalities was rated the worst ($M = 3.38$, $SD = 1.54$). A one-way ANOVA did not reveal a significant difference for the combination of modalities ($p > .05$). The effect size estimate shows a large effect ($\eta^2 = .220$).

In the interviews participants stated that *"[they] generally like the haptic feedback that is provided by the glove."* (P5, P13). However, they also stated that *"the haptic feedback has a little more delay than the other modalities"* (P8, P9 ,P10, P16). Considering the auditory feedback P9 stated that *"sound is a nice way of communicating an error, however it is very unpleasant when people are standing next to me. They will immediately know that I made an error."* Considering the visual feedback, participants liked that it is displayed very fast (P1, P13). All feedback modalities combined were considered as *"too much feedback."* (P3).

9.4 Summary

In this chapter, we evaluated haptic, auditory, and visual error feedback modalities for assistive systems at a manual assembly workplace. Through two lab studies, we first found suitable variants for each modality, and second, identified how the feedback modalities are perceived by participants.

The results show that participants considered the auditory error feedback, although only noticeable when standing at the workplace, to significantly violate the workers' privacy. Considering the general rating, the participants rated the visual and haptic error feedback similarly well. Auditory error feedback was rated worse than the haptic and visual error feedback. Further, auditory error feedback was rated as more distracting compared to haptic error feedback that is provided by a glove equipped with vibration motors, or visual feedback that is given using a projector. Our results further indicate that combining haptic and visual feedback might be a good choice for communicating errors at the workplace. Although combining visual and haptic error feedback did not significantly differ from the other combinations, the quantitative results and the qualitative feedback indicate that using haptic error feedback combined with visual error feedback to communicate errors might be appropriate.

Chapter 10

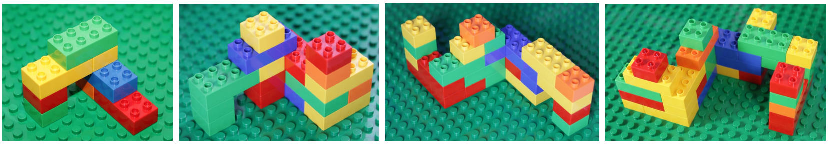
Evaluation of Instruction Creation Methods

This chapter is based on the following publication:

- M. Funk, L. Lischke, S. Mayer, A. Sahami Shirazi, and A. Schmidt. Assistive augmentation at the workplace: A system for creating semantically-rich assembly instructions. In *Assistive Augmentation: Cognitive Science and Technology (to appear)*. Springer, 2017

In the previous chapters, we found that different instructions for communicating work steps exist. However, as storing products is expensive and ordered products can be highly customized, we envision that in the near future many companies will produce with a lot size of one. Hence, this means that each produced part is different from the previously produced part, the cognitive effort at the workplace will increase. Thus, task-specific instructions are becoming more important to cognitively support the workers at the workplace. In this chapter, we want to address the research question: How can in-situ instructions for assistive systems be created? (RQ4)

For creating task-specific instructions, two types of instructions can be easily created without requiring technical knowledge: First, recording video instructions



(a) The 8 brick model (b) The 16 brick model (c) The 24 brick model (d) The 32 brick model

Figure 10.1: The constructions used in the lab study with four different complexity levels: (a) 8 bricks, (b) 16 bricks, (c) 24 bricks, and (d) 32 bricks.

using a camera and second, using Programming by Demonstration (PbD) (cf. Section 4.4) to create in-situ projected instructions. As the choice of the appropriate instruction is also dependent on the quality of the instruction and the time it takes to create them, we conduct three studies. First, we conduct a lab study evaluating time, errors, and cognitive effort that result from using either type of instruction with different complexities. In a second study, we measure the time and cognitive effort that it takes to create either type of instruction using ten experts from the industry by conducting the study in an industrial setting. We show that the creation effort of in-situ instructions is comparable to the creation effort of video instructions, where no semantic information is retained. In a third study, we show that using the instructions that were created by experienced workers can be used in real production environments for assembling real products.

10.1 Study#1: Different Complexities

To assess assembly tasks with a different complexity and a different number of work steps, we conducted a user study in our laboratory. Inspired by previous work [149], we decided to use Lego Duplo[22] bricks for creating constructions with different numbers of bricks. As our test setup consisted of eight bins, we considered four models with four different numbers of bricks, i.e., 8, 16, 24, and 32. All the four models were created using eight different types of bricks in five different colors. They all have one arch in the bottom level. Figure 10.1 shows the four constructions.

[22] `http://www.lego.com/en-us/duplo` (last access 5th Oct. 2016)

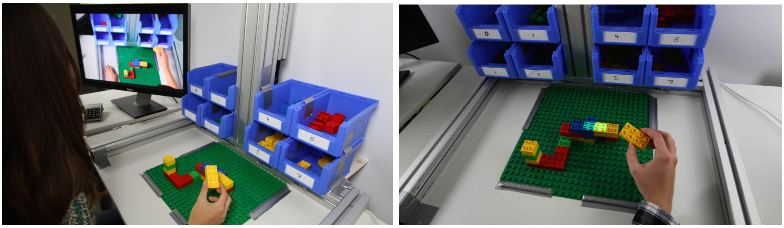
10.1.1 Method

A mixed design was considered for carrying out this study. We used a between-subject design with the type of instruction as the only independent variable with two levels: the video-based approach and the in-situ projection approach. Within the groups, we used a repeated measures design with the number of bricks as the independent variable (four levels). As dependent variables in both groups, we measured the ER, the TCT, and the NASA-TLX score. The order of the repeated measures tasks was counterbalanced according to the Balanced Latin Square.

We created two assembly instructions for each construction model: recording the video, and the PbD approach. For recording the video instructions, we used a camcorder and videotaped the assembly instructions in HD resolution recorded over the shoulder of the worker. For recording the projected instructions, we used our PbD system. In both cases one of the researchers performed the assembly task while the instructions were recorded and created. For both conditions, the content of the bins and their arrangement was identical. Each type of brick had a separate bin resulting in eight different bins.

For the video instruction, a monitor was placed next to the work area (see Figure 10.2(a)). The participant could play and pause the video using the space key on the keyboard at any time during the assembly. For the in-situ projection, the participant sat in the same place in front of the plate and for each step instructions were projected into the work area by either highlighting a bin to pick from or projecting the contour on the position where the brick should be placed (see Figure 10.2(b)).

The procedure of the study was as follows: after welcoming the participant and giving a brief introduction about assembling products, we collected the demographics. Then, one of the instructions was assigned to the participant. When the participant was ready, the experimenter started the instruction and measured the TCT. The participant was instructed to only use the predominant hand to pick and assemble the bricks. The whole experiment session including hands and the picking from the bins was video recorded for each participant. After assembling each model, the participant was asked to fill in the NASA-TLX questionnaire. The participant repeated this procedure for all four construction models. After the study, two researchers independently watched the videos and counted the errors for each participant. They compared the results and in case of inconsistency, the researchers reviewed the videos together till they came to an agreement.



(a) Participant using video instruction (b) Participant using in-situ instruction

Figure 10.2: The setup of the two conditions used in the lab study. (a) The video condition uses a monitor to display the instructions. (b) The in-situ condition projects visual feedback onto the Duplo bricks.

We recruited 32 participants, 8 female and 24 male that were aged from 20 to 38 years ($M = 25.1$ years, $SD = 3.9$ years) using the University's mailing list. All participants were students in various majors. They had no prior knowledge in assembling the Duplo constructions nor participated in the two previous studies. Furthermore, none of the participants was colorblind. The study was conducted in our lab at the University of Stuttgart. The light conditions were constant during the whole study.

10.1.2 Results

We statistically compared the ER, the TCT, and the NASA-TLX score between the four models and the two instruction methods conducting a two-way mixed ANOVA. Mauchly's test indicated that the assumption of sphericity had been violated for ER ($\chi^2(5) = 17.60$, $p < .004$) and TCT ($\chi^2(5) = 23.29$, $p < .001$). Therefore, the degrees of freedom were corrected using Greenhouse-Geisser correction ($\varepsilon = .73$ for ER and $\varepsilon = .68$ for TCT). The t-test with Bonferroni correction was considered as post hoc test for all cases.

The analysis revealed that the difference in the ER between the four models was not significant ($F(2.18, 65.36) = 1.94$, $p > .05$). The model with the 24 steps had the largest ER ($M = .66$, SD 1.61) followed by the 32-step model ($M = .59$, $SD = 1.38$) and 16-step model ($M = .47$, $SD = 1.04$). Whereas, the effect on the ER between the two feedback approaches was statistically significant ($F(1, 30) = 11.20$, $p < .002$, $r = .39$). The effect size estimate shows a medium and hence substantial effect. The post hoc test showed that the video-based instruction had a significantly larger ER than the in-situ projection instruction (M = .86, SD = 1.36 vs. M = .05, SD = 1.36, $p < .002$).

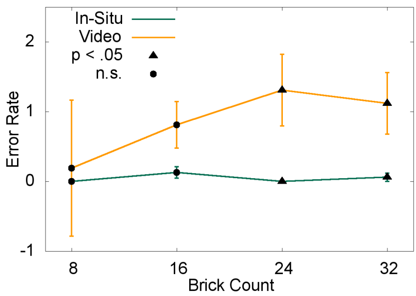
Analyzing the TCT between the construction models showed that it differed statistically significantly ($F(2.05, 61.5) = 217.88$, $p < .001$). Post hoc tests revealed a significant difference between all constructions. The 32-step model had the longest TCT (M = 2.31 minutes, SD = .69) followed by 24-step (M = 1.83 minutes, SD = .70) and 16-step (M = 1.10 minutes, SD = .31). Such differences were already expected due to the variation in the number of bricks. However, the feedback approaches had a statistically significant effect on the TCT ($F(1,30) = 63.82$, $p < .001$, $r = .80$). The effect size indicates a large and substantial effect. Surprisingly, the TCT using the video method took 1.5 times longer than the PbD method (M = 1.73 minutes, SD = .45 vs. M = 1.08 minutes, SD = .45).

Furthermore, there was a statistical significant difference in the NASA-TLX score between the construction models ($F(3,90) = 3.63$, $p < .01$). The post hoc tests showed that the difference was only significant between the 8-step and 32-step models (M = 22.34, SD = 16.20 vs. M = 27.87, SD = 17, $p < 0.1$). The score between other constructions was not significant (all $p > .05$). The average score for the 16-step model was 25.03 (SD = 17.47) and for the 24-step construction the score was 26.38 (SD = 16.93). The comparison between the methods revealed a statistically significant effect on the perceived cognitive load ($F(1,30) = 19.73$, $p < .001$, $r = .54$). The effect size indicates the effect is large and substantial. The perceived cognitive load for the in-situ instruction approach was 60% smaller than the video-based instruction (M = 15.62, SD = 25.96 vs. M = 35.19, SD = 25.96, respectively).

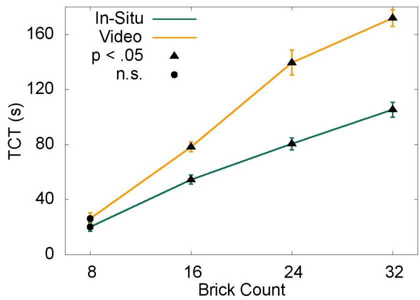
Impacts of number of steps in assembly

We further assessed the differences between the two feedback approaches having different numbers of assembly steps, respectively, different complexities. To achieve this, for each construction model, we conducted the t-test between the video and in-situ instructions and pair-wisely compared the ER, TCT, and NASA-TLX score. The Levene's test was conducted in all cases to test the equality of variances. In case the assumption was violated the degrees of freedom were adjusted.

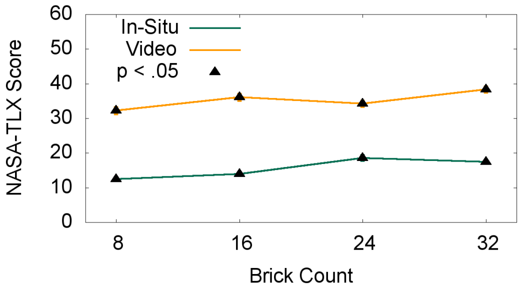
The comparison of the ER showed the in-situ instruction had fewer errors than the video instruction in all levels of complexity (see Figure 10.3(a)). The difference was not significant in the 8-step ($t(15) = 1.86$, $p > .05$, $r = .43$) and 16-step



(a) Error rate



(b) Task completion time



(c) Nasa TLX score

Figure 10.3: The results of the lab study for constructions with different number of steps: (a) error rate (ER), (b) task completion time (TCT), and (c) NASA-Task Load Index (NASA-TLX) score.

constructions ($t(16.84) = 1.94$, $p > .05$, $r = .42$). However, the difference was statistically significant in the 24-step construction ($t(15) = 2.48$, $p < .05$, $r = .53$) and the 32-step construction ($t(15) = 2.31$, $p < .05$, $r = .50$). The effect size estimate indicates that the effect on the ER for all four models using the provided instructions is large.

The comparison of TCTs revealed that the difference between both approaches was statistically significant for all steps except for the 8-step construction ($t(30) = 1.11$, $p > .05$, $r = .20$). Figure 10.3(b) shows the average TCT for the four construction models using the two instruction methods. In all cases the TCT was significantly faster using the in-situ approach (for the 16-step montage: $t(30) = 4.69$, $p < .001$, $r = .65$; for 24-step montage: $t(30) = 5.67$, $p < .001$, $r = .71$; for 32-step montage: $t(30) = 7.92$, $p < .001$, $r = .68$). The effect sizes show that the effect of the provided instructions on the TCT of the assembly tasks is large except for the 8-step assembly task.

Further, the NASA-TLX scores statistically significantly differ in all four construction models (see Figure 10.3(c)). In all cases the score for the in-situ instruction was significantly lower than the video approach: for the 8-step montage, $t(19.37) = 4.30$, $p < .001$, $r = .70$; for 16-step montage, $t(20.52) = 4.58$, $p < .001$, $r = .71$; for 24-step montage, $t(30) = 2.90$, $p < .007$, $r = .47$; for 32-step montage: $t(30) = 4.37$, $p < .001$, $r = .62$. The effect size estimate indicates that the effect on the perceived cognitive load using the two instruction approaches is large, and therefore substantial for all models.

10.1.3 Discussion

The results of the user study reveal that there are significant differences between the video instruction and the in-situ projection approach in the ER, the TCT, and the perceived cognitive load during the assembly tasks. Using the in-situ instruction, the ER decreases up to 17%, the TCT is up to 1.5 times faster, and the perceived cognitive load is reduced up to 60% in comparison to the video-based instruction.

Further, the comparison of the in-situ and video-based instructions in different levels of complexity unveil that the in-situ instruction outperforms the video-based approach independent of the number of steps. In all levels the ER is lower and the TCT is faster. These differences are statistically significant when the number of steps in the assembly task increases. Moreover, the perceived cognitive load is significantly lower for the in-situ instruction independent from the number of steps in the assembly task.

10.2 Study #2: Creating Instructions

To evaluate our system for creating assembly instructions in a real assembly scenario, we conducted a user study using a real assembly task (a refurbished car's engine starter) with industrial workers. We made this conscious choice to increase the validity of the results, even if it is harder to reproduce the results. Using students and a lab-based study is in our opinion not appropriate to address these questions.

10.2.1 Method

We use a repeated measures design with three conditions for creating an instruction: by demonstration using our assistive system, using a graphical editor, and video recording. The only independent variable is the creating-method. As dependent variables, we measure the task completion time (TCT) and the NASA-TLX score [77]. The order of the conditions is counterbalanced.

For the condition using the graphical editor, we re-implemented the system presented by Korn et al. [98] as a control condition. In contrast to our assistive system, the user should use a GUI to manually highlight the bins, the workpiece carrier, or tools that have to be used for the assembly task using different geometric shapes (see Figure 10.4(a)). Further, the GUI is used to define actions in each step of the assembly and create an instruction. For the video condition, we recorded a video of the assembly from the worker's point of view. A camcorder was installed behind the user in such a way that the worker's point of view could be simulated. The participant had to inform the experimenter when the video recording should be started and stopped.

As the assembly task for the study, we chose the assembly of a car's engine starter (see Figure 10.4(c)). The task consists of five steps and in each step one part has to be assembled. When all five parts are put together on the workpiece carrier, the worker has to assemble two screws on top of the starter using a screwdriver tool.

We carried out the study in a car manufacturing company in Germany. After welcoming the participants and explaining the course of the study, we collected the demographics. Next, we introduced the participants to the workpiece carrier and let them get familiar with it. We allowed the participants to assemble the engine starter twice to get themselves familiar before starting the study. Afterwards, the study was started and participants had to create instructions using

(a) GUI of the editor (b) Real world editing (c) Engine starter used in the study

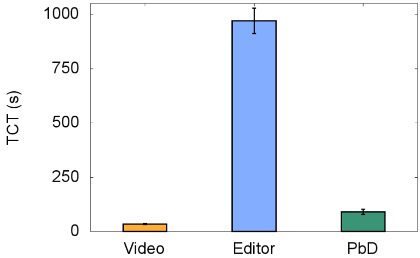
Figure 10.4: (a) The graphical editor allows changing the properties of projected elements. (b) The worker can adjust the projection directly on the workpiece carrier. (c) We are using a car's engine starter as assembled product in the user study as a real assembly scenario.

the three approaches: editor, PbD, and video. At the end of each condition, the experimenter measured the TCT. Afterwards, the participant completed the NASA-TLX questionnaire. At the end, we collected qualitative feedback through semi-structured interviews.

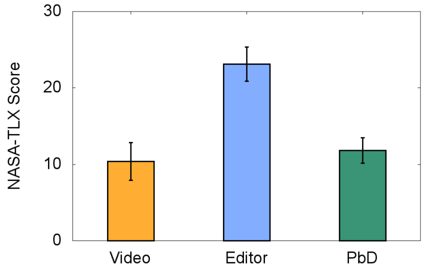
We recruited 10 workers from the company (2 female, 8 male), who were familiar with the engine starter. The participants were aged between 17 and 53 years ($M = 32.1$, $SD = 13.9$). All participants had experience in assembling the engine starter for at least one year and can be considered as experts.

10.2.2 Results

We statistically compared the TCT between the instruction creation methods. Mauchly's test indicated that the assumption of sphericity had been violated ($\chi^2(2) = 18.04$, $p < .0001$). Therefore, the degrees of freedom were corrected using the Greenhouse-Geisser correction ($\varepsilon = .52$). A repeated measures ANOVA showed that the TCT statistically significantly differs between the methods ($F(1.05, 9.49) = 256.04$, $p < .0001$, $r = .97$). The effect size estimate reveals a large and therefore substantial effect. Post hoc tests using Bonforroni correction revealed a significant difference between all three methods (all $p < .05$). The



(a) Task completion time



(b) NASA-TLX score

Figure 10.5: (a) The average task completion time that was needed for creating the instructions using each method. (b) The average NASA-TLX score that was scored when creating instructions using each method. The error bars depict the standard error.

video method had the shortest TCT ($M = 0.58$ minutes, $SD = .08$) followed by the PbD ($M = 1.52$ minutes, $SD = .63$) and the editor ($M = 16.16$ minutes, $SD = 3.07$). Figure 10.5(a) shows a graphical representation of the results.

Further, we statistically compared the NASA-TLX scores between the instruction creation methods. The sphericity assumption was not violated ($p > .05$). A repeated measures ANOVA determined that the methods used had a statistically significant effect on the NASA-TLX score ($F(2,18) = 19.83$, $p < .0001$, $r = .81$). The effect size estimate shows a large and substantial effect. Post hoc tests using the Bonferroni correction revealed that the editor had a significantly higher perceived cognitive load ($M = 23.10$, $SD = 7.79$, $p < .007$) than PbD ($M = 11.80$, $SD = 5.22$) and the video ($M = 10.40$, $SD = 7.07$, $p < .001$). However, the difference between PbD and the video was not statistically significant ($p > .05$). The results are shown graphically in Figure 10.5(b).

The qualitative feedback showed that the participants found the editor hard to use. Although they were experts in assembling an engine starter, they didn't have enough experience in using a computer (e.g., P6, P1). Further, a participant stated that *"using the editor is too time-consuming"* (P3). One participant had also privacy concerns when recording a video as co-workers could identify him based on his hands and his wristwatch (P4).

10.2.3 Discussion

The results of the study reveal that the editor approach induces a significantly higher perceived cognitive load in comparison to the PbD and video approaches. Whereas, there is no significant difference in mental demand required for creating assembly instructions using the PbD and video approaches. Hence, the additional perceived cognitive load added due to the use of our interactive system is not significant.

The results further show that recording the video is the fastest way for creating an assembly instruction followed by PbD and the editor approach. One reason is that no additional time is required to capture the depth information after each assembly step. In contrast, the PbD approach requires that the users shortly remove their arms and head from the work area to capture the depth data of the product.

Although the PbD-based and video-based approaches are faster and require less mental effort than the editor-based approach in creating instructions, the approaches might differ when it to comes to assemble the engine starter using the created instructions. Therefore, we conduct a follow-up study to evaluate the instructions while assembling the engine starter with novice users using the created instructions.

10.3 Study #3: Using Instructions

In the previous study we assessed the PbD approach for creating assembly instructions. In order to evaluate the practicality of the created assembly instructions in assembling a real product, we conduct a followup study assembling the same engine starter, which we used in the previous study.

10.3.1 Method

For providing assembly instructions, we use the instructions created in the previous study. We randomly choose one instruction created using each approach resulting in three different instructions for assembling the engine starter: (1) the video-based assembly instruction, (2) the in-situ projection instruction created using the editor, (3) the in-situ projection instruction created using the PbD approach. For the in-situ projection instruction using the editor, the user explicitly created the instruction using a graphical editor. In contrast, our system automatically generated the other instruction. We chose a between subject design with three groups to prevent a learning effect between the different instructions. The only independent variable that differed between the groups was how the instruction was created. As dependent variables we measured the number of errors (ER), the task completion time (TCT), and the NASA-TLX score.

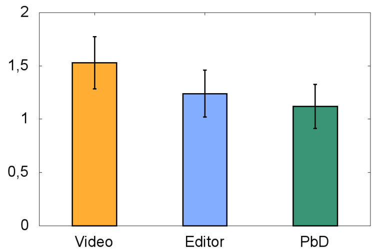
We conducted the study in the same car manufacturing company as in the previous study. After welcoming the participant and explaining the course of the study, we collected the demographics and ensured that the participant never assembled an engine starter before. Then, the participant was accompanied to our prototype and one of the instructions was assigned and explained. As the participants did not differ in skills, the condition was randomly assigned. The participant was told to assemble an engine starter based on the instructions provided. When the participant was ready, the experimenter started the instruction and counted the ER. The TCT was measured by the system automatically. During the assembly the experimenter did not provide any help. After the assembly was done, the participant was asked to fill in a NASA-TLX questionnaire. Finally, qualitative feedback was collected through semi-structured interviews.

We recruited 51 participants (12 female, 39 male) aged between 23 and 60 years ($M = 47.8$, $SD = 9.3$). We divided the participants equally between the conditions, resulting in 17 participants per condition. All participants were employees of the car manufacturing company and were unfamiliar with the assembly task and the product, i.e., assembling an engine starter. Hence, they can be considered novice users. None of the recruited participants took part in the previous study.

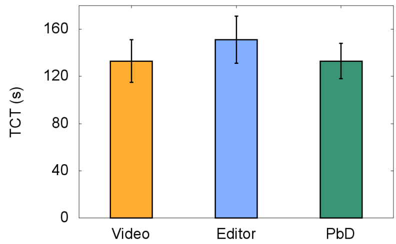
10.3.2 Results & Discussion

We statistically compared the ER, the TCT and NASA-TLX between the groups. The assumption of homogeneity of variance had not been violated ($p > .05$). A one-way ANOVA test revealed no statistical significant effect on ER between the groups ($F(2,48) = .89$, $p > .05$). The group using the instruction created by our PbD system had the lowest ER ($M = 1.12$, $SD = .86$), followed by the group using the instruction created by the editor ($M = 1.24$, $SD = .90$), and the group using the video-based instruction ($M = 1.53$, $SD = 1.01$). The results are depicted in Figure 10.6(a).

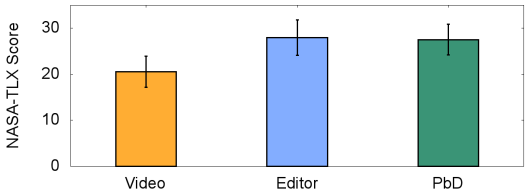
The statistical analysis also revealed no significant difference in the TCT between the groups ($F(2,48) = .32$, $p > .05$). The group using the instruction created by



(a) Error rate



(b) Task completion time



(c) NASA-TLX score

Figure 10.6: (a) The average error rate that were made using each type of instruction. (b) The average TCT that was needed for using each type of instruction. (c) The average NASA-TLX score that was scored when using each type of instruction. The error bars depict the standard error.

the PbD approach had the shortest TCT ($M = 2.21$ minutes, $SD = 1.05$) and the group using the instruction created with the editor had the longest TCT ($M = 2.52$ minutes, $SD = 1.39$). The group using the video instruction took on average 2.22 minutes ($SD = 1.31$) to assemble the engine starter. A graphical representation of the results is shown in Figure 10.6(b).

The analysis showed no statistical significant effect on the NASA-TLX score between the groups ($F(2,48) = 1.38$, $p > .05$). The group using the video-based instruction had the lowest perceived cognitive load ($M = 20.59$, $SD = 13.90$), followed by the group using the instruction created using the PbD approach ($M = 27.53$, $SD = 13.87$), and the group using the instructions created by the editor ($M = 28$, $SD = 15.84$). Figure 10.6(c) provides a graphical overview.

The qualitative feedback indicated that the in-situ projected instructions were generally well perceived. The participants particularly found the step by step feedback of the projected instructions very helpful (P42, P33). Additionally, they mentioned that directly projected feedback onto the workplace was very useful (P30, P12). One participant stated that *"I don't have to think anymore while working"* (P24). Another participant mentioned that *"I would rather work autonomous in the daily life, but for training I would use it"* (P45). Participants using the video instruction mentioned that the video was helpful for learning the task instead of having an instructor (P22, P51) but they didn't want to work when a video is playing all day (P37, P25).

The analysis shows that the ER is reduced and the TCT is faster in the assembly task using instructions that were created using the PbD approach compared to the other two approaches. However, the change is not statistically significant. The results indicate that the instruction automatically created using our system slightly performs better than the explicitly created instruction using the editor. Moreover, the results suggest that the in-situ projection increases the perceived cognitive load during the assembly, but the change is not significant. The qualitative feedback indicates that the step by step instructions provided directly in the work area through the in-situ projection is more accepted than a video-based instruction.

As the assembly task only consists of five steps, no big differences were expected. Based on the results from study#1, it can be expected that with more steps the differences between these instructions would increase and a clearer advantage for instructions that were created using PbD would be revealed.

10.4 Implications

The previously described user studies revealed implications on both creating assembly instructions and performing an assembly task based on previously created instructions. In the following we discuss the insights gained through these studies.

10.4.1 Creating Assembly Instructions

The results of the studies indicate that creating assembly instructions using the PbD approach is as intuitive as recording a video. Creating instructions using the editor approach is more time-consuming and increases the perceived cognitive load compared using the PbD approach and the video-based approach.

Using the PbD approach, the time required to create an instruction is higher than recording a video of the assembly as our system requires that the user waits 1.5 seconds between each step to detect that a step was performed. The time is even higher when using the editor approach as the user has to manually specify each step. However, editing steps in both PbD and editor approach is easier than editing video-based instruction since each step can be modified separately. In contrast, video-based instructions need to be post-processed and manually edited. Editing videos can be complex and may result in re-recording the video even if only a single step needs to be altered. Another advantage of using either editor-based or PbD-based instructions is that they additionally store depth information of each step. Using the depth information the assistive system can monitor if the correct part is picked and if the part was correctly assembled.

While the video approach induces the lowest perceived cognitive load when creating instructions, the results further show that using in-situ instructions does not significantly increase the perceived cognitive load. However, the editor approach induces a higher perceived cognitive load by interacting with the GUI.

A further advantage of including semantic information into the instructions is that the instructions can be transferred to a specific work place automatically. One could even imagine that an experienced worker in one company (or one country) can create the instructions remotely and these instructions can then be downloaded to an assembly workplace in another company (or country) (cf. [61]).

10.4.2 Assembly Performance

When it comes to assembling a product, the results suggest that the in-situ projection approach reduces the perceived cognitive load of the worker, the TCT, and the ER compared to a video-based instruction. Especially, these effects are significant when the number of assembly steps increase. As the in-situ assembly instructions are provided directly in the work area, the distraction is minimized compared to showing the videos on a monitor close to the work area. This reduces the perceived cognitive load that is required for following instructions and also reduces the TCT for assembling a product. Furthermore, our assistive system's step by step error control can monitor if the correct part is picked and if it is assembled correctly using depth information. This leads to fewer errors even with increasing number of work steps.

10.5 Summary

In this chapter, we evaluated the concept of using PbD to create semantically-rich assembly instruction. Using PbD for creating in-situ instructions is faster than using a graphical editor. In contrast to just recording video, which is slightly faster, in-situ instructions retain all features of interactive instructions and do not add any significant perceived cognitive load to the user in comparison to assembling with the help of the video instructions. The instruction creation approaches were evaluated with experienced workers in a production environment using a real product. In a large laboratory study, we showed that in-situ instructions outperform the video-based instruction in assembly tasks with different numbers of steps. In-situ instructions decrease the error rate, the task completion time, and the perceived cognitive load. This trend was also observed when assembling a real product in a production environment, however not statistically significant.

Chapter 11

Evaluation of Long-Term Impact

As our assistive system for the workplace is running very robustly and the system is able to detect parts that are larger than 1cm$\times$1cm$\times$1cm with a high accuracy, using the system over a longer period of time becomes possible. Therefore, we are able to conduct a study lasting for more than one week per participant, in which we evaluate the long-term effects of using our assistive system for assembly tasks for a whole work day. Simultaneously, we evaluate the effectiveness of our envisioned concept for adaptive assembly feedback (cf. Section 5.1.2). With this study, we are able to address our research question considering the long-term effects of using in-situ instructions (RQ5). As the different user groups involved in manual assembly have different requirements for assistive systems, we split the population into three user groups as outlined in Section 2.2: experienced workers, inexperienced workers, and workers with cognitive impairments. We measure the effects of in-situ projected instructions for each user group in a separate study. In this thesis, we focus on the evaluation of in-situ instructions for experienced workers and inexperienced workers. Throughout the study with these user groups, we produced 2145 car's engine starters in a total assembly time of 56.8 hours using our in-situ instructions provided by our assistive system. The results reveal that using our assistive system for more than one week resulted in a decrease of performance for experienced workers. However, inexperienced workers learned the assembly steps just by following the instructions provided by the system.

INFO: This thesis encompasses the long-term evaluation for experienced and inexperienced workers. As a part of the motionEAP project, we additionally evaluated the long-term effects of assistive systems on workers with impairments. This study is described in detail in the PhD thesis of Liane Bächler (cf. [12]).

11.1 A Long-Term Study with Experienced and Inexperienced Workers

For evaluating the long-term effects of in-situ instructions provided by our assistive system on experienced workers and inexperienced workers, we conducted a study at a major car manufacturing company in Germany. We installed our assembly cell (cf. Section 5.4.2) at one of their factory buildings and conducted two experiments following the same experiment design. The only difference between the two experiments is the experience of the participants: either experienced workers or inexperienced workers.

11.1.1 Design

We designed the study according to a repeated measures design with the in-situ instructions as only independent variable with two levels (with in-situ instructions and without in-situ instructions). As dependent variables, we were measuring the TCT, the number of errors, and the NASA-TLX [77] score. We did not counterbalance the order of the conditions as we were interested in the initial learning effect of the in-situ instructions on the workers. As the assembly cell is a U-shaped assembly line, where all three workplaces work together on one product, the errors and TCT are measured for the entire group of three workers.

11.1.2 Apparatus

For the study, we used the U-shaped assembly cell (cf. Section 5.4.2) that is equipped with our assistive system (see Figure 11.1). The assembly cell is used to assemble cars' engine starters and was deployed at a major car manufacturing company. We especially created the assembly cell for being able to use the assistive system on the whole assembly process of the car's engine starter. The way of

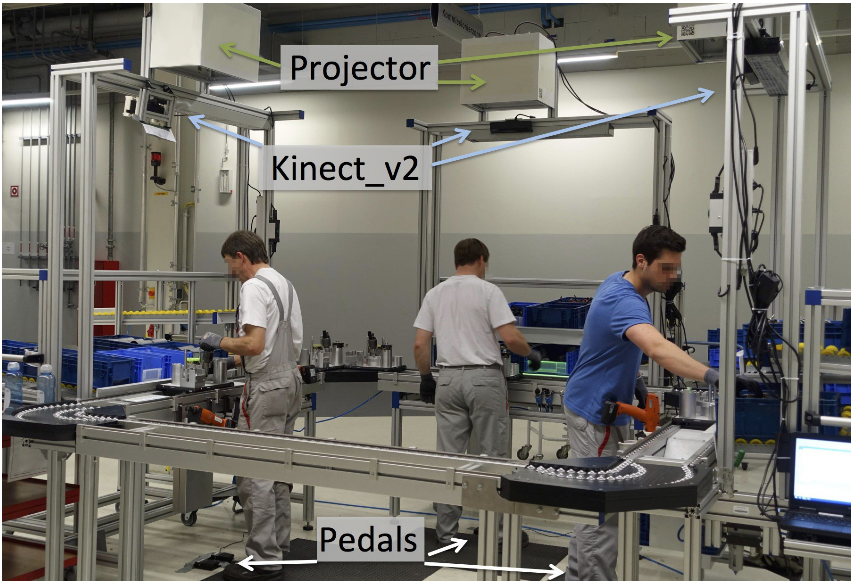


Figure 11.1: We installed our assistive system for providing in-situ assembly instructions using the assembly cell setup. Each workplace is equipped with a Kinect_v2 and a projector. The work steps involving assembly parts that are too small to detect with the camera can be advanced manually using foot pedals.

assembling the engine starters differs from the usual assembly process that the car manufacturing company uses to assemble the engine starters. We deliberately introduced a new apparatus to create a new setting for both experienced workers and inexperienced workers. Further, we used a new workpiece carrier, which holds the assembled product during assembly. It was designed in a way that every performed work step can be seen by the top-mounted Kinect_v2. For the study, we used seven workpiece carriers that can be transferred between the workplaces along the U-shaped assembly cell (see Figure 11.1). The transferring of the workpiece carriers has to be done manually. In the study, the workpiece carriers were transferred between the workplaces in a counterclockwise direction. At each workplace a pneumatic clamp was firmly mounting the workpiece carrier at the exact same position that the Kinect_v2 was able to perform the assembly detection. As the assembly also involved parts that are smaller than $1\text{cm}\times 1\text{cm}\times 1\text{cm}$, there were assembly steps that could not be detected by the assistive system due to camera limitations. Therefore, we used a foot pedal with two buttons for advancing the steps manually. One button advances the workflow by one step and the other button can be used to switch to the previous step.

As the workplaces are dependent on each other, we split the work steps that are performed at each workplace in a way that all workers need the same amount of time for performing work steps at each workplace. For workplace 1 (WP1), the task consisted of 18 work steps: twelve could be automatically detected, six had to be advanced manually because the parts were too small. Workplace 2 (WP2) consisted of twelve work steps, thereby three had to be advanced manually and nine were detected automatically. Finally, workplace 3 (WP3) consisted of fourteen work steps: four had to be advanced manually and ten were detected automatically. However, if one worker needs more time, the next worker in line needs to wait for the previous workplace to finish. To introduce a buffer in case a workplace needs more time than usual, we use seven workpiece carriers at the assembly line.

Considering the software used in this study, for presenting instructions we use three adaptivity levels (cf. Section 6.4): beginner level, advanced level, and expert level. Inline with the findings from Chapter 8, we use a contour information for highlighting the assembly position of the picked parts using a green light. In the beginner level, this contour information is extended with instruction videos taken from the user's point of view for learning how to perform the work step. In the advanced level, just the contour information is shown. Finally, in expert level, the system does not show any information. However, if an error occurs, the system displays red error feedback (cf. Chapter 9), e.g. when picking from a wrong bin. As described in Section 6.4, we use the error ratio as a trigger for switching between the adaptivity levels. In this study, the thresholds were set to 0.05 for switching to expert level and 0.15 for switching to medium level. These thresholds differ from the thresholds that we used in the single workplace lab setting, as there were more picking errors made in the assembly cell setup. In case an error was made, the system immediately switched back to the previous adaptivity level for the work step where the error was made. The succeeding work step was again using the error ratio as criteria for the adaptivity level.

11.1.3 Procedure

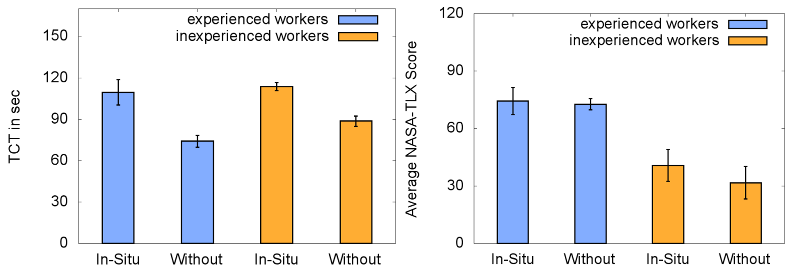
As we conducted the study at a car manufacturing company, we had to stick to the company's breaks and hours of work. Thus, a workday consisted of four slots with a total of 360 minutes of assembling, which is exactly 6 hours per day. Another 1.5 hours were needed per day to prepare the study for the next day, evaluating the study after the assembly time, and for collecting feedback from the participants. As the workers should learn the entire assembly process on each of the three workplaces (WP1, WP2, WP3), all workers had to work on each of the three assembly workplaces in the assembly cell. We iterated the workplaces in a counterclockwise way after each break, resulting in iterating four times per workday. As the assembly cell is built in a U-shape, after the final step was performed at WP3, a study assistant removed the assembled engine starter from the workpiece carrier, put it into a container, and counted the produced starters. Afterwards, the assistant moved the empty workpiece carrier in reach of WP1 again. The errors were counted in a post-assembly quality control. Thereby, a quality inspector checked the engine starters for assembly errors. After three days of assembling, we conducted a group interview, where we invited the participants and asked them for their opinion about using in-situ instructions at the workplace. To measure the learning effect, we afterwards assembled the car's engine starter using the same assembly cell for three days without using instructions. We started the procedure with the group of inexperienced workers and then repeated the same procedure for the experienced workers. However, due to time limitations of the car manufacturing company, we had only two instead of three days for the condition assembling without instructions for the group of experienced workers. This results in an overall study run time of eleven workdays.

11.1.4 Participants

We recruited 3 experienced workers and 3 inexperienced workers (all male), who all are employees of a major car manufacturing company. The experienced workers were on average 43.34 years old ($SD = 4.49$ years) and the inexperienced workers were on average 45.67 years old ($SD = 12.65$ years). All experienced workers had at least one year of experience in assembling the engine starter. The inexperienced workers had experience in working on assembly tasks but did not assemble an engine starter before. All workers were not familiar with the workpiece carrier and the U-shaped assembly cell as it was especially designed for this experiment. Further, all participants volunteered to take part in the study.

11.1.5 Results

We report the results of the study for both experienced workers and inexperienced workers separately. For both user groups, we statistically compared the TCT,



(a) Average TCT

(b) Average NASA-TLX score

Figure 11.2: (a) The average time per produced part for experienced workers and inexperienced workers using in-situ instructions and assembling without instructions. (b) The average NASA-TLX score using in-situ instructions and assembling without instructions. All error bars depict the standard error.

number of errors, and NASA-TLX between the two conditions using a one-way repeated measures ANOVA.

Experienced Workers

Considering the TCT, the experienced workers were faster without instructions with an average assembly time of 74.03s ($SD = 11.63$s) per produced part compared to during the learning phase using in-situ instructions, which resulted in an average assembly time of 109.40s ($SD = 31.96$). The ANOVA revealed a significant difference between the approaches $F(1,6) = 8.428$, $p = .027$. The effect size shows a large effect ($\eta^2 = .584$). A graphical representation is depicted in Figure 11.2(a).

When comparing the NASA-TLX between the two conditions, the average NASA-TLX score for the experienced workers was 74.34 ($SD = 12.25$) using the in-situ instructions and 72.67 ($SD = 4.98$) without instructions. A statistical comparison using a one-way repeated measures ANOVA test did not reveal a significant difference between the conditions ($p > .05$). A graphical representation of the average TLX scores is depicted in Figure 11.2(b).

As the assembly errors that were made during the study were determined in a post-process, only descriptive statistics can be reported for the number of errors. However, the experienced workers did not make any assembly errors both with and without the in-situ instructions.

Inexperienced Workers

When considering the average time to produce a part, the inexperienced workers were faster after the learning phase without using instructions with an average assembly time of 88.65s (*SD* = 12.41s) compared to 113.62s (*SD* = 10.14s) during the learning phase using in-situ instructions. The one-way repeated measures ANOVA revealed a significant difference between the conditions $F(1,10) = 23.621$, $p = .001$. The effect size estimate shows a large effect ($\eta^2 = .703$). The results are depicted in Figure 11.2(a).

The post-process analysis of errors revealed that the group of inexperienced workers made five errors while working with the in-situ instructions. They did not make any assembly errors afterwards without using the instructions.

We further compared the NASA-TLX scores between the two conditions for the inexperienced workers. Using the in-situ instructions, the questionnaire resulted in an average score of 40.67 (*SD* = 14.34) and an average score of 31.67 (*SD* = 14.70) without using in-situ instructions. A statistical comparison using a one-way repeated measures ANOVA did not reveal a significant difference between the conditions ($p > .05$). A graphical representation of the average TLX scores is depicted in Figure 11.2(b).

Between User-Groups

Considering experienced workers and inexperienced workers as different user groups, we statistically compare the results as a between groups experiment. We use a one-way repeated measures ANOVA to compare the TCT, number of errors, and NASA-TLX score when using in-situ instructions and assembling without instructions afterwards.

When comparing the TCT for assembling the engine starter while using in-situ instructions, the ANOVA did not find a significant difference between the experienced workers and the inexperienced workers ($p > .05$). However, when comparing the TCT for assembling the engine starter without using instructions, the ANOVA found that the experienced workers were significantly faster than the inexperienced workers, $F(1,6) = 6.589$, $p = .043$. The effect size estimate shows a large effect ($\eta^2 = .523$).

We further compared the average NASA-TLX score between the experienced workers and inexperienced workers. However, the ANOVA test did not reveal a significant difference for using the in-situ instructions ($p > .05$) and for assembling without instructions ($p > .05$).

11.1.6 Qualitative Results

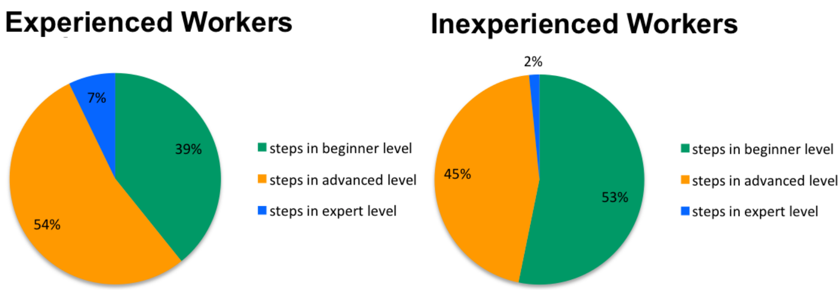
For better understanding the effects of the in-situ instructions on both groups of workers, we provide detailed results from the interviews that were conducted after each condition. Thereby P1-P3 are from the inexperienced workers group and P4-P6 belong to the group of experienced workers.

Participants disliked that the workplaces that were used in the study differed from the workplaces they were used to. Especially that the assembly cell constructed for the study was designed in a U-shape as *“[they] had to wait for the previous worker to finish”* (P2). However, after working with the in-situ instructions for a longer time, a participant stated that *“[he] got used to the system”* (P1). One participant told us that *“working with the in-situ instructions was relaxing”* (P3). In contrast, participants perceived the in-situ instructions as *“an additional task, which required [them] to pay extra attention to the colors”* (i.e. not causing red error feedback). Sometimes participants also were *“irritated because of the triggered red light”* (P5) that indicated a wrong pick. Using the in-situ instructions while working was perceived as *“working twice as much”* (P2) because *“manually advancing the instructions with the foot pedal lead to a higher physical load on the left foot”* (P6). However, the workers liked that they were supported during their tasks because *“the system shows us how the order of the work steps has to be performed”* (P1). However, one participant stated that *“[he] felt like a robot”* (P6). Overall, most participants stated that *“[they] wouldn’t want to work with the system every day”* but it *“would be great to learn new tasks with the system.”*

11.1.7 Results of Adaptive Visual Feedback

Considering the adaptive feedback, we measured how many steps were performed in which adaptivity level. Thereby, we only counted the adaptivity level that was active when the work step was successfully performed. In the study, the experienced workers performed 25349 work steps, whereas the inexperienced workers performed 26409 work steps. It has to be mentioned again that the study of the experienced workers was one day shorter than the study of the inexperienced workers. As the study run time and the produced parts is different, we chose to only report the results descriptively.

Considering the experienced workers, the most steps were performed in the medium level ($n = 13584$) followed by the beginner level ($n = 9943$) and the



(a) Adaptivity levels used by experienced workers (b) Adaptivity levels used by inexperienced workers

Figure 11.3: The distribution of work steps that were performed in each adaptivity level (a) considering experienced workers and (b) considering inexperienced workers.

expert level ($n = 1822$). A distribution of the used levels is depicted in Figure 11.3(a). In contrast, the inexperienced workers performed the most steps in the beginner level ($n = 14052$), followed by the medium level ($n = 11938$). The expert level was only used in 419 work steps. A graphical representation of the used adaptivity levels for the inexperienced workers can be found in Figure 11.3(b).

11.2 Discussion

The results reveal that both experienced workers and inexperienced workers are significantly faster in assembling the car's engine starter without in-situ instructions than assembling with in-situ instructions. Considering the perceived cognitive load using the NASA-TLX score, we did not find a significant difference between the conditions for both experienced workers and inexperienced workers. While experienced workers did not make any mistake, the inexperienced workers only made mistakes during the in-situ projection condition. However, as we did not counterbalance the order of the conditions, we cannot draw a conclusion. The qualitative feedback indicated that in the first day, the participants enjoyed learning how to assemble the engine starter using the in-situ projection. Further, the participants stated that the study run time of assembling three days using in-situ projected instructions was too long. After they have learned the assembly steps and felt confident in assembling the product, they would have liked to turn off the projection.

Considering the adaptivity levels, we were initially expecting that after the participants learned the assembly steps, the system will switch to expert level, turn off the in-situ instructions, and only will intervene if a participant is about to make an error. However, the inexperienced workers only finished 2%, and the experienced workers only finished 7% of all work steps in the expert level. After the study, we identified two circumstances as reasons for the adaptivity levels to not switch into expert level. First, we found that the system detected a lot of picking errors when the participants picked from a bin that was located at the left side using their right arm. This caused the trigger-boxes that were defined between the participants arms and the target bin to trigger. Consequently, the falsely triggered boxes were counted as picking-errors which causes the error ratio to increase and the system to switch to a lower adaptivity level. As a second reason for the system to not stay in the expert level after the participant has learned the work steps, might be the design of the visual feedback in the expert level. Through observations and interviews, we found that the participants were confused when changing from advanced level into expert level as the projection was turned off after they completed the work step. P2 told us that *"[he] thought that the system had crashed because there was no projection shown anymore."* As we did not show any feedback in expert level to not distract the participants, we unintentionally violated one of Shneiderman's golden rules [145], which states that the user should always be in charge of the system and know when a state is changed. With automatically switching the adaptivity level to expert level without notifying the users, we caused confusion which lead to users staying at the medium level.

11.3 Summary

In this chapter, we evaluated the long-term impact of in-situ projected instructions on experienced workers and inexperienced workers. In a controlled field study with a total run time of eleven workdays, we deployed our assembly cell prototype in an assembly hall of a major car manufacturing company. We were able to observe two effects and reveal one weakness of the system. As effects, we found that inexperienced workers are able to learn assembly steps using in-situ projected instructions. While the group of inexperienced workers made five assembly errors during the initial learning phase, after learning the assembly steps using the in-situ projection, they were able to assemble the engine starter without the help of instructions and without making any errors. As a second effect, we observed that continuously projecting in-situ instructions at the assembly area slows down both experienced workers and inexperienced workers significantly. Thereby, it has to be considered that we did not counterbalance the order of the conditions to observe the feasibility of learning assembly steps using in-situ instructions. Qualitative results revealed that the duration of presenting in-situ instructions was perceived as too long by the workers. They stated that learning the steps one day instead of three days would have been sufficient. Thus, we assume that in scenarios with inexperienced and experienced workers, in-situ projection is only suitable for training scenarios rather than continuously supporting workers during their tasks.

As another outcome of the long-term study, we observed that our design of the adaptivity levels and our model of classifying adaptivity levels based on picking errors was not suitable for a long-term usage. As our assistive system unintentionally triggered picking steps caused by natural movements of the workers, there is a need for using other classifiers for switching between adaptivity levels. One approach could be using physiological data of the workers [48].

Chapter 12

Comparing Modalities for Order Picking

This chapter is based on the following publications:

- M. Funk, A. S. Shirazi, S. Mayer, L. Lischke, and A. Schmidt. Pick from here!: An interactive mobile cart using in-situ projection for order picking. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 601–609. ACM, 2015

- M. Funk, S. Mayer, M. Nistor, and A. Schmidt. Mobile in-situ pick-by-vision: Order picking support using a projector helmet. In *Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, New York, NY, USA, 2016. ACM

In this chapter, we evaluate the two order picking systems that were introduced in Chapter 7 against state-of-the-art picking approaches. First, we compare the OrderPickAR picking cart against Pick-by-Paper, Pick-by-Voice, and Pick-by-Vision. Second, we compare the body-worn HelmetPickAR order picking system against a state-of-the-art Pick-by-Paper approach. As the two studies were

Figure 12.1: The layout of the warehouse that was used in the study. Each compartment is labeled with a compartment number. The warehouse consists of 30 different compartments.

conducted independently from each other, we describe the design and setup in two separate sections. With this chapter, we are able to answer our research question considering how in-situ instructions can be used for order-picking tasks (RQ6).

12.1 The OrderPickAR Picking Cart

In the following, we describe the setting of the study comparing the OrderPickAR picking cart and three other picking methods that we derived from related work and industrial applications. Further, we describe the study design and the results of the user study.

12.1.1 Warehouse Layout

For conducting our user study, we built a warehouse consisting of three shelves in our research lab. Figure 12.1 shows the layout of the warehouse. The shelves are aligned to form a grid consisting of 5 rows $\times$ 6 columns. Each compartment is labeled with an identifying number. In total, the warehouse is 2.07m high and 3.62m wide. We designed the warehouse in a way that the distance to walk between the picks is minimal. However, the field of view of the shelf-facing camera-project pair is approximately half of the size of the warehouse depending on the distance of the cart to the shelves. The position of the compartments in the warehouse is not ordered according to features of the stored items. Each compartment contains ten items of the same type. We did not use bins inside the compartments and stored the items directly in the shelves. As items to pick, we use 30 different Lego bricks in different shapes and colors. This warehouse layout is designed to represent a classical picker-to-parts low-level warehouse [37].

12.1.2 Picking Methods

In the following, we describe the picking methods that we used in the evaluation of the OrderPickAR prototype. In addition to our OrderPickAR picking cart, we considered three other existing approaches: Pick-by-Paper (PbP), Pick-by-Vision (PbVi), and Pick-by-Voice (PbVo).

Pick by Paper (PbP)

The PbP approach (Figure 12.2(a)), where a worker gets a paper list containing the picking information, is still used in many warehouses in industry. As other research [70, 143, 152] uses PbP as a method to compare a new system against, we include PbP as a baseline in our study. We are using a paper list containing the following information about each picking task: article's description, article's number, quantity to pick, source compartment, and destination bin. The user has to find the source compartment in the warehouse, pick the correct quantity, and place the items into the destination bin.

Pick by Voice (PbVo)

For the PbVo approach, we recorded audio instructions for all picking tasks. The user can control the audio instruction with the following commands: *next*, *back*, and *replay*. We designed the commands to reflect state of the art systems[23]. The command *next* jumps to the next (or first) instruction and plays it back. The *back* command jumps to the previous instruction and plays it back. The *replay* command replays the current instruction again without proceeding to the next instruction. The instructions were played back in a headset (see Figure 12.2(b)). As we did not implement a voice input, a wizard of oz issued the playback of the picking instruction according to the participants commands. The audio instructions contained the information needed for the current picking task, e.g., *"pick 3 items from shelf 02-10 and put them into bin 39"*. In contrast to state-of-the-art systems, to make the system more comparable to the other systems used in the study, we decided not to include a *ready* command to confirm the pick.

Pick by Vision (PbVi)

To further compare our OrderPickAR picking cart to existing approaches, we implemented a PbVi system using the attention funnel visualization [23] on an Epson Moverio BT-200[24] HMD. This approach is similar to the approach of Schwerdtfeger et al. [143]. The attention funnel visualization displays circles towards the compartment to pick from (see Figure 12.2(c)). Further, it displays the quantity of the items that have to be picked in the bottom left corner of the HMD. For tracking the position and orientation of the HMD, we equipped it with OptiTrack markers at each side. The position and orientation information is calculated at a desktop computer and transmitted to the HMD via WiFi. Further, the HMD is running a 3D visualization of the attention funnel using the Unity3D engine. The program then uses the position and orientation information of the HMD and the position of the target compartment to adjust the visualization. The position and orientation information is received with 100 frames per second. However, due to the limited processing power of the Moverio BT-200, our PbVi system was only able to display 9-10 frames per second. The system is implemented in a way that it only renders the most recent position information. Position frames that cannot be processed in time are dropped.

[23] `http://www.dematic.com/en/Supply-Chain-Solutions/By-Technology/Voice-and-Light-Systems/Pick-to-Voice` (last access 5th Oct. 2016)

[24] `http://www.epson.com/moverio` (last access 5th Oct. 2016)

(a) The Pick-by-Paper method (b) The Pick-by-Voice method

(c) The Pick-by-Vision method (d) The OrderPickAR system

Figure 12.2: The picking methods that were used in our study. (a) A participant using the Pick-by-Paper list. (b) The Pick-by-Voice approach. (c) Perspective of the participant using Pick-by-Vision with an attention funnel visualization. (d) The projector cart highlights the bin to pick from.

12.1.3 Method

For the study, we used a 4-level repeated measures design with the guidance system as the only independent variable. The guidance systems used were: PbVi, PbVo, PbP, and the OrderPickAR projector cart. As dependent variables, we measured the error rate (ER), the task completion time (TCT), and the NASA-TLX [77] score.

After welcoming the participant and explaining the course of the study, a general introduction about order picking was given. Before each condition, the participant was allowed to perform three picking tasks to get familiar with the current guidance system and moving the picking cart. The picking task we are using in the study is the so-called *discrete picking* [40]. This means that the participant takes parts from one compartment and puts it into a single bin on the cart afterwards. After placing the picked items the user picks from another compartment.

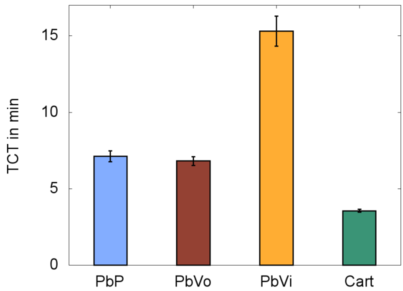
We considered four picking tasks each consisting of 10 different steps and 30 items to pick. The items were located in different compartments of the previously described warehouse. The participants were told to always move the cart in front of the shelf, where they need to pick the next part from. The movement of the cart was needed to simulate a regular-sized warehouse. Furthermore, we chose to only use the shelf-facing side of the cart as it provided enough bins. We designed all picking tasks to require the same distance that needs to be walked between the shelves and the cart. Further, the distance that the cart has to be moved is the same in each task. We counterbalanced the order of the conditions and tasks using the Balanced Latin Square. The participants were also instructed to carry all items belonging to a step in one single walk from the shelf to the cart and not to split the picking task into multiple walks. As this study focuses on the different types of feedback, we designed the study that all feedback is proceeded by a wizard of oz during all conditions. Further, the facilitator counted the ER. After completing one task using one condition, the participant was asked to fill in a NASA-TLX [77] questionnaire. We repeated the procedure for all conditions. At the end, we collected additional qualitative feedback.

We instructed all participants to focus on not making any errors during the picking tasks, which is considered the primary goal. Further, we told the participants that a fast picking of the orders is only considered the secondary goal, nevertheless the time to pick the orders is measured during the study.

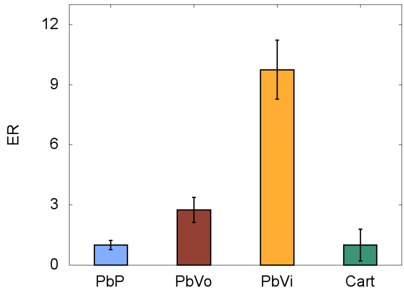
We recruited 16 participants (4 female, 12 male) via our university's mailing list. The participants were aged from 20 to 43 years ($M = 24.81$, $SD = 5.39$) and were students with various majors and a secretary. None of the participants had experience with order picking. All participants were not familiar with our system or the picking tasks. The study took approximately 60 minutes per participant. The participants were compensated with 5€.

12.1.4 Results

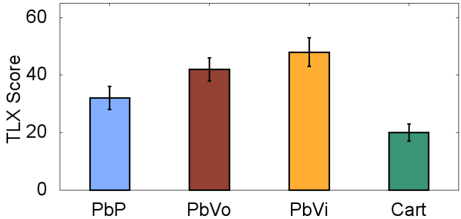
We statistically compared TCT, ER, and NASA-TLX score between the guidance systems using a one-way repeated measures ANOVA. Mauchly's test showed that the sphericity assumption was violated for TCT($\chi^2(5) = 37.70$, $p < .001$)



(a) Task completion time



(b) Number of errors



(c) The NASA-TLX score

Figure 12.3: The results of our study: (a) task completion time in minutes, (b) error rate, and (c) the mental load indicated by the NASA-TLX score.

and ER ($\chi^2(5) = 18.16$, $p < .003$). Therefore, we used the Greenhouse-Geisser correction to adjust the degrees of freedom ($\varepsilon = .42$ for TCT and $\varepsilon = .57$ for ER). Otherwise stated, the Bonferroni correction was used for all the post-hoc tests.

A repeated measures ANOVA showed a statistically significant difference in TCT between the approaches, $F(1.287, 19.305) = 93.99$, $p < .001$. The post-hoc tests depicted that the differences between all approaches are significant (all $p < .05$) except between PbVo and PbP ($p = n.s.$). Figure 12.3(a) shows the average TCT of all approaches. The TCT was fastest using the OrderPickAR system ($M = 3.55$ minutes, $SD = 0.38$) followed by PbVo ($M = 6.81$ minutes, $SD = 1.12$), PbP ($M = 7.11$ minutes, $SD = 1.40$), and PbVi ($M = 15.31$ minutes, $SD = 3.89$).

The statistical analysis also revealed a significant difference in the ER between the approaches, $F(1.711, 25.667) = 22.49$, $p < .001$. The post hoc test only showed a significant difference (all $p < .05$) between PbVi and all other approaches. The OrderPickAR projector cart ($M = 1$, $SD = 3.24$) and PbP ($M = 1$, $SD = .96$) had the lowest ER, followed by PbVo ($M = 2.75$, $SD = 2.56$) and PbVi ($M = 9.75$, $SD = 6.07$). The results are shown in Figure 12.3(b).

The analysis reveals a significant effect on the NASA-TLX scores between the approaches, $F(3,45) = 11.06$, $p < .001$. The post-hoc test only revealed a significant difference (all $p < .05$) between OrderPickAR and all other approaches (see Figure 12.3(c)). OrderPickAR had the lowest score ($M = 20.25$, $SD = 12.64$) followed by PbP ($M = 32.25$, $SD = 17.48$), PbVo ($M = 42.44$, $SD = 17.30$), and PbVi ($M = 48.81$, $SD = 21.00$).

The qualitative feedback indicated that participants did not find the visualization provided on the HMD helpful. They mentioned that the glasses slightly moved when performing a pick, which caused the funnel to become inaccurate (P10, P14). Further (P11) mentioned *“[I] would not like the head-mounted display when working together with co-workers"*. The participants found the in-situ feedback provided by OrderPickAR fast and easy to use (P10). But they sometimes occluded the projection when picking from lower bins (P7). Overall, participants liked that they had both hands free for performing the picking task in the PbVi, PbVo, and projector cart conditions. They disliked that they had to carry the picking list at all times in the PbP condition.

12.1.5 Discussion

The results suggest that using our system has several advantages. First, the TCT is almost 2 times shorter than using the PbVo and PbP approaches and even more than 4 times faster than using the PbVi approach. Second, the ER is significantly lower (up to 9 times compared to the HMD approach). The difference in number of errors in comparison to the classical paper-based approach was not significant. Interestingly, all errors that were made with the OrderPickAR approach in the user study were made by one single user, who did not pay attention to the displayed number of items to pick. Third, the mental demand during the order picking is more than two times lower than using other interactive approaches. The qualitative feedback also conveys that users find the in-situ feedback projected directly on the shelves and bins better than using an HMD. The feedback on the HMD may hinder users to communicate with each other. As the HMD slightly moves during a picking task, the displayed visualization was introducing an offset.

When comparing our results to previous work using the same PbVi representation, our PbVi approach performs worse compared to the respective PbP approach.

Schwerdtfeger et al. [143] and Reif et al. [133] report a similar TCT comparing PbP and PbVi, while our PbVi approach differs significantly from our PbP approach. Concerning ER, previous work reported a slightly higher rate when using the PbVi approach compared to PbP [143]. However, our PbVi approach performs significantly worse compared to our PbP approach. This difference might be caused by the used HMD, the Epson Moverio BT-200. Especially, as the qualitative feedback of the participants revealed that, although fitting the HMD to each user in a calibration step, the viewing accuracy in the PbVi system was prone to fast head movements. Therefore, the HMD needed to be corrected after the HMD was moved before continuing. The participants mentioned that the HMD is relatively heavy compared to normal glasses. Also the HMD's rubber parts behind the participants' ears and on the participants' noses sometimes caused the HMD to slip to the bottom of the participants' noses. We can imagine that using a different HMD for the study, the results of the PbVi system would be different.

12.1.6 Limitations

Despite the fact that our system could support multiple users working on a picking task, we limited the task to only support a single user as this is favored by industry. A multi-user scenario could, e.g., use color coding to assign targets to users or use the top-mounted Kinects to track different users. Further, the task used in the evaluation of the system was limited to 10 steps. Tasks with a different amount of steps may reveal other results. In the study, we were using a wizard of oz approach to advance the feedback in case the pick was not registered correctly. E.g., if the user is occluding the Kinect's field of view with the head, the user would have to manually advance the feedback in an industry setting. Additionally, the OrderPickAR approach is only able to detect that the user picked from the correct compartment, however it is not able to detect how many items were picked from it. This problem is solved in industry settings by adding a scale into the process. Especially for larger quantities, the order's weight is checked in an additional step. For determining the indoor position of the cart in our proof-of-concept system, we use the OptiTrack motion capturing system. Using this technology is not feasible in larger warehouses because the cost of covering larger areas is too high. To scale this up to larger warehouses and to reduce the costs of the system, an approach using visual markers for determining the position and orientation could be used in future versions of the system. E.g., Kim et al. [89] use an optical marker-based solution for tracking the indoor position of the user. A similar approach could also be used for tracking the picking cart in a large warehouse.

12.2 The HelmetPickAR Projector Helmet

We further evaluate the HelmetPickAR system by conducting a user study which compares the HelmetPickAR system to the state-of-the-art Pick-by-Paper instruction method.

12.2.1 Design

We conducted a repeated measures study with the used order picking instruction system (HelmetPickAR or Pick-by-Paper) as the only independent variable. As dependent variables, we measure the task completion time, the number of errors, and the NASA-TLX (Task Load Index) score. We use two different tasks for the study. The tasks consist of the same number of items to pick and the same walking distance that is required for the worker to perform the task. To be able to analyze the performance for each picking step, we measure the task completion time for both picking and placing the items separately. We further collected qualitative feedback through semi-structured interviews after the study. Finally, we asked for the participants' subjective ranking of the used order picking instruction systems. The order of the conditions and the used tasks were counterbalanced.

12.2.2 Apparatus and Warehouse Layout

For evaluating the HelmetPickAR system (which is described in detail in Chapter 7), we considered using the state-of-the-art Pick-by-Paper as a control condition. We created our PbP method according to the PbP methods that were described in previous research (e.g. [70]). The PbP method consists of a paper list that is arranged using a table. Each order is described in one row of the table. The columns of the table contain the following information: ID, article number, amount to pick, source compartment, destination position, price of the item, and a free column that can be used as a checkbox.

As a warehouse, we were using the same 5×6 compartments that we used in the previous OrderPickAR study. Like in the previous experiment, the compartments can be identified by a number that is written on a label directly at the compartment. Each compartment contains Lego bricks in a way that each compartment holds ten bricks having the same color and shape. We additionally drew a white line that was 20cm in front of the shelves to prevent the participants from crashing into the shelves with the helmet. The target positions were different from the OrderPickAR study, as we were using a 120cm×60cm table that was divided into three equally large target positions using white lines. The table with the target positions was positioned 2m away from the shelves. Using this layout having a complex warehouse with easy target positions, we can observe both, effects on rather easy tasks, and the effects on more complex tasks.

12.2.3 Procedure

After welcoming the participant, we explained the course of the study and gave a general introduction about order picking. We informed the participants about their right to quit the study at all times and asked them for permission to take pictures during the study. After signing a consent form, we collected the demographic information. Then, we gave an introduction to the first order picking instruction method according to the counterbalanced order of the conditions. Considering the PbP method, we did not instruct the participants to use the checkboxes but told them that they could use them to check already processed orders. Further, we did not tell the participants to take the picking list with them while processing the orders. They could also leave the list at one place, however, this would increase the task completion time. Considering the HelmetPickAR condition, we first showed a picture to the participants containing all used symbols and explained their meaning (cf. Figure 7.3(b)). Further, we firmly mounted HelmetPickAR on the participant's head using a chinstrap and gave them about two minutes to get used to the helmet. The participant was given two example orders to practice the current order picking instruction method. Once the participant felt confident in using the instruction method, and had no further questions, we started with the first picking task using the instruction method. We instructed participants that the first priority was to not make any errors and that the second priority was to process the orders fast. After performing the picking task, we asked the participants to fill in a NASA-TLX [77] questionnaire. As errors, we counted picking from a wrong compartment, placing the picked items at a wrong target position, and picking the wrong number of items. We repeated the procedure for the remaining task and instruction method. After finishing the second task, we asked the participant to rank the instruction methods according to their subjective preference. Finally, we collected additional qualitative feedback through semi-structured interviews.

12.2.4 Participants

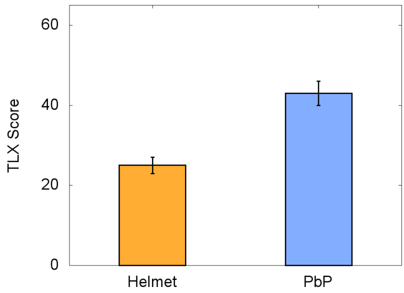
For the study, we recruited 16 participants (6 female, 10 male) via our university's mailing list. The participants were aged from 19 to 37 years (M = 23.8, SD = 6.34). Thirteen of the participants were students with various majors and three were employed in industry. None of the participants was familiar with the picking tasks or order picking in general. All participants were using HelmetPickAR for the first time. The study took approximately 40 minutes per participant. The participants were compensated with candies.

12.2.5 Results

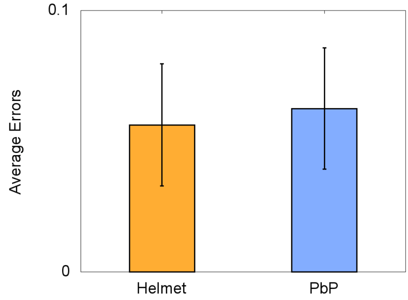
We statistically analyzed the TCT divided into picking time and placing time, the number of errors, and the NASA-TLX score between HelmetPickAR system and the PbP instruction using a one-way repeated measures ANOVA.

First, we analyzed the picking time between the two order picking instruction systems. The participants on average needed 12.6 seconds (SD = 2.96 seconds) to perform a pick using HelmetPickAR and an average of 13.53 seconds (SD = 2.7 seconds) using the PbP instruction. The one-way repeated measures ANOVA did not reveal a significant difference for the picking time, $F(1,15) = 1.703$, $p > .05$. The effect size estimate shows a medium effect ($\eta^2 = .102$). Considering the time that a participant needed to place the picked order at a target position, the participants on average needed 20.45 seconds (SD = 4.03 seconds) using HelmetPickAR and 17.51 seconds (SD = 3.62 seconds) using the paper baseline. A one-way repeated measures ANOVA revealed a significant difference between the two order picking instruction systems considering placing time, $F(1,15) = 8.183$, $p = .012$. The effect size estimate shows a large effect ($\eta^2 = .353$). Figure 12.4(c) depicts the differences between the approaches and the picking times.

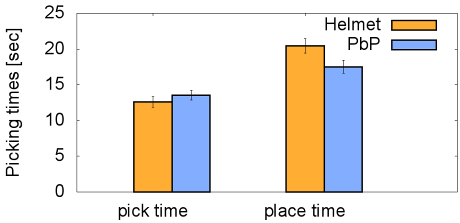
Regarding the number of errors that the participants made using both instruction systems, the participants made on average 0.056 errors per order (SD = 0.096 errors per order) using HelmetPickAR and 0.063 errors per order (SD = 0.096 errors per order) using the PbP instruction (see Figure 12.4(b)). The one-way repeated measures ANOVA did not reveal a significant difference for the number of errors, $F(1,15) = 0.028$, $p > .05$. The effect size estimate shows a small effect ($\eta^2 = .002$).



(a) The NASA-TLX score



(b) Average errors



(c) Task completion time

Figure 12.4: The results of the user study comparing the HelmetPickAR system and the Pick-by-Paper baseline: (a) the average cognitive load using the NASA-TLX score, (b) the average number of errors that was made during a picking task, and (c) the task completion time for both picking and placing tasks. The error bars depict the standard error.

Considering the perceived cognitive load according to the NASA-TLX questionnaire using both systems, the participants rated HelmetPickAR with an average score of 25.37 ($SD = 11.25$) as less cognitively demanding compared to the PbP baseline instruction with an average score of 43.38 ($SD = 15.66$). Figure 12.4(a) also depicts the scores for both conditions. The one-way repeated measures ANOVA revealed a significant difference between the two order picking instruction systems considering the NASA-TLX score, $F(1,15) = 17.060$, $p < .001$. The effect size estimate shows a large effect ($\eta^2 = .532$).

For the subjective ranking after finishing both conditions, 12 participants preferred HelmetPickAR and 4 participants preferred the PbP baseline. In the interviews, two participants stated that they *"enjoyed being guided by the order picking helmet, because [they] did not have to search for the correct compartment and* *just follow the arrows."* (P3, P4). However, one participant thought that using HelmetPickAR is *"just following instructions blindly without having to put in cognitive effort."* (P12). They further stated that when performing the order picking tasks *"[they] felt like robots."* (P7, P12, P15). However, in general participants liked that *"the head-mounted projection is always visible and in the line of sight."* (P3). Considering the paper baseline, a participant stated that *"carrying a paper picking list interferes with the picking task."* (P1).

12.3 Summary

In this chapter, we evaluated the two proposed order picking systems OrderPickAR and HelmetPickAR by conducting two user studies comparing the systems to state-of-the-art instruction systems. Considering the evaluation of the OrderPickAR system, we compared OrderPickAR to three existing approaches, i.e., Pick-by-Vision, Pick-by-Voice, and Pick-by-Paper. The evaluation shows that the OrderPickAR projector cart is faster than the other approaches and significantly reduces the user's perceived cognitive load. Additionally, we found that the OrderPickAR approach reduces the number of errors compared to a Pick-by-Vision approach. The users find the in-situ feedback provided directly on the warehouse and on the shelves more helpful than feedback on an HMD or just having a picking list. For evaluating the HelmetPickAR, we conducted a user study with 16 participants, where we compared HelmetPickAR against the state-of-the-art Pick-by-Paper approach. We found that there is no significant difference between HelmetPickAR and PbP in errors made and picking time, however PbP is significantly faster than HelmetPickAR considering placing time. We assume that this is due to the simple design of the target positions. A more complex target position design might yield different results. Lastly, we found that the perceived cognitive load is significantly lower using HelmetPickAR compared to the PbP approach. Overall participants liked the experience of having an augmented view of the world without having to wear a HMD.

Chapter 13

A Benchmark for Instruction Giving

This chapter is based on the following publications:

- M. Funk, T. Kosch, S. W. Greenwald, and A. Schmidt. A benchmark for interactive augmented reality instructions for assembly tasks. In *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia*. ACM, 2015
- M. Funk, T. Kosch, and A. Schmidt. Interactive worker assistance: Comparing the effects of head-mounted displays, in-situ projection, tablet, and paper instructions. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2016

After evaluating the assistive system that was proposed in this thesis and carefully reviewing related work about assistive systems, we found that optimizing these systems in a scientific way is very difficult. The challenge is to introduce performance metrics that apply across different tasks and find a uniform experiment design to evaluate proposed instruction systems in a comparable way. Other areas in the field of Human-Computer Interaction have already recognized the need for standardized tasks. For example, for evaluating text-entry techniques, the phrase set of MacKenzie et al. [110] is considered the standardized task. In this chapter, we address this challenge by proposing a standardized experiment design for evaluating interactive instructions and making them comparable with each other. Further, we introduce a General Assembly Task Model (GATM), which differentiates between *task-dependent* and *task-independent* measures. Through a user study with twelve participants, we evaluate the experiment design and the proposed task model using an abstract pick-and-place task and an artificial industrial task. Further, we provide paper-based instructions for the proposed task as a baseline for evaluating AR instructions and comparing them with state of the art AR instructions. With this, we address our research question considering how instructions for workplaces can be evaluated (RQ7). In the second part of this chapter, we use the GATM method and introduced reference tasks to evaluate our assistive system and compare it to other state-of-the-art instructions: HMD instructions, tablet instructions, and baseline paper instructions. Our results show that assembling parts is significantly faster using in-situ projection and locating positions is significantly slower using HMDs. Further, participants make less errors and have a lower perceived cognitive load using in-situ instructions compared to the HMD instructions.

13.1 A General Assembly Task Model

All research projects mentioned in Chapter 3 provide ways of giving instructions to workers. However, each presented approach uses its own assembly task. This makes the different instructions very difficult to compare against each other, as each new assembly task introduces a different complexity and different times to assemble a product. Unfortunately, most research only reports the total time that workers need to assemble a product, which is only a *task-dependent* measure. This problem in comparing different instruction approaches with one another could be solved by finding *task-independent* measures or by standardizing experiments.

According to the literature, a standardized approach for planning how long a worker will need for different steps of an assembly task is the MTM (methods-time measurement) [113] approach. Thereby, each single movement of a worker is assigned to groups of pre-defined standard movements, which have a standard time for performing the movement. This helps planners in industrial planning scenarios to estimate the time that workers will need for assembly steps. In this work, we want to use the concepts behind MTM, generalize them according to our manual assembly workplace scenario, and analyze which of the actions can be treated as *task-independent* measures.

In order to find *task-independent* measures, we analyzed the tasks that were used in related work and constructed an equation which can be applied to each presented task. As related work identified the four relevant component actions for assembly tasks to be *reach, grasp, move position, and release* [25], we abstract the basic actions to high-level actions that require cognitive effort and are affected by an instruction. Thus, we suggest the General Assembly Task Model (see Figure 13.1), which calculates the total time a worker needs to assemble a product t_{total} by measuring four sub-times (t_{locate_part}, t_{pick}, t_{locate_pos}, and $t_{assemble_x}$).

$$t_{total} = n\left(t_{locate\_part} + t_{pick} + t_{locate\_pos} + t_{assemble_x}\right)$$

Figure 13.1: The equation for calculating the assembly time according to the General Assembly Task Model.

Thereby, n is the number of assembly steps required. The time to locate the bin in which the next part is stored is t_{locate_part}. This includes both the cognitive time to process the instruction and the time to move the arm to the target bin. The latter can be treated as a constant value as workplaces are usually designed in a way that the distance to pick parts is within an arms reach [155]. The time that a worker needs to pick a part is t_{pick}. If the specific task does not include picking a part, t_{locate_part} and t_{pick} are 0. t_{locate_pos} is the time that the worker needs to locate the assembly position of the part that the worker is currently holding or the part that the worker needs to modify. Finally, $t_{assemble_x}$ is the time needed by the worker to perform the assembly task associated with a specific part x.

task-dependent measures: $t_{assemble_x}$ is *task-dependent* since different parts might take more or less time to assemble. It is also a measure for instruction quality, as instructions can communicate how to use tools during an assembly step, how to correctly assemble a part, and which details to focus on.
t_{pick} is a *task-dependent* measure, as it depends on the size and weight of the part that needs to be grasped. However, we don't consider t_{pick} to be a measure for instruction quality since grasping is usually not instructed.
task-independent measures: We consider t_{locate_part} and t_{locate_pos} as indicators for the *task-independent* instruction quality, as they quantify the cognitive effort that the worker needs to transfer the given instructions to the workplace.

13.2 Requirements for a Standardized Task

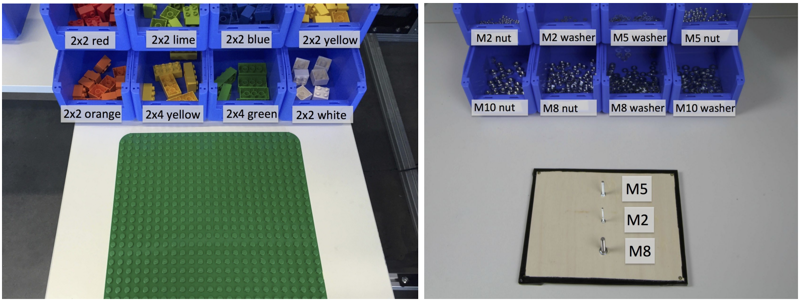
In order to find a standardized assembly task for instruction giving, we analyzed tasks that were used in previous work and identified two categories of tasks. Besides our own experiments, other related approaches [14, 138, 149] recognized Lego Duplo as abstraction for industrial pick-and-place tasks. The major benefit of such abstract pick-and-place tasks is that the time that a worker needs is mainly t_{locate_part} and t_{locate_pos}, where $t_{assemble_x}$ is relatively low, as the brick is just placed at a position and no further assembly is required. However, related work uses specific industrial assembly tasks [32, 81, 165, 166]. Compared to a pick-and-place task, $t_{assemble_x}$ is much higher when using industrial assembly tasks. As the most time is used to perform the assembly itself, the t_{locate_part} and t_{locate_pos} only account for a small part of t_{total}. Overall, we recognize two different types of tasks, pick-and-place tasks and industrial assembly tasks. Thereby, each work step belonging to either type consists of one or many of the following three actions: picking items, placing items, and assembling them.

To sum up the requirements for a standardized task for comparing assembly instructions with one another, we define the following four criteria for designing a uniform assembly task:

- **cheap to setup:** the proposed task has to consist of off-the-shelf components that are affordable and easy to purchase.
- **easy to replicate:** the assembly tasks have to be described in a way that they are easily replicable.
- **representative:** the tasks have to cover the three main actions that can be found in assembly scenarios in the industry (i.e. picking parts, placing parts, and assembling parts).
- **easy to scale up:** the number of work steps have to be easily changeable without changing the nature of the task.

13.3 A Uniform Experiment Design

Inspired by approaches from prior work, we propose a uniform experiment design as a benchmark for assembly instructions that fulfills the requirements mentioned



(a) Duplo task (b) Industrial assembly task

Figure 13.2: The apparatus of the two proposed reference tasks. (a) The Lego Duplo setup, where the used bricks are stored in blue picking bins. They are arranged as depicted. The assembly plate is directly in front of the picking bins. (b) An industrial assembly task using screws and nuts. The assembly parts are stored in blue picking bins and the workpiece carrier is placed in front of the bins. The screws are aligned vertically that the worker can use both hands to for assembling the nuts and washers.

above. We implement two assembly tasks which follow the proposed design: A pick-and-place task using Lego Duplo, and an industrial task that requires a precise assembly of components (see Figure 13.2).

The benchmark follows a repeated measures design with the number of assembly steps as the only independent variable. We consider tasks with 4, 8, 16, and 32 assembly steps for both tasks. As dependent variables, we suggest measuring the assembly time according to the GATM (t_{locate_part} ,t_{pick} ,t_{locate_pos}, $t_{assemble_x}$), the errors that were made during assembly, and the perceived cognitive load using the NASA-TLX [77] questionnaire. As the time and errors are dependent on the number of assembly steps, we normalize all times and errors by dividing them by the number of assembly steps.

In order to measure the exact assembly times, the experiment has to be recorded and evaluated in a post-analysis. The time t_{locate_part} is measured from the moment the instruction is shown until the hand of the worker is in the correct bin. Second, t_{pick} is the time the participant's hand is inside the bin. Next, t_{locate_pos} is the time from when the hand leaves the bin until the part arrives in the right location for the assembly task. Finally, $t_{assemble_x}$ is the time at the appropriate site to perform the assembly task. Further, the order of the number of assembly steps should be counterbalanced according to the Balanced Latin Square.

13.3.1 Duplo Bricks

The first task is an abstract pick-and-place task using Lego Duplo bricks. For the task, we equip a workplace with a 26×26 green Lego Duplo plate (see Figure 13.2(a)). The Duplo bricks that are used for the assembly are stored in eight picking bins that are arranged in a 2×4 grid. The exact arrangement and content of the bins is depicted in Figure 13.2(a). We use eight different Lego Duplo bricks (size 2×2: orange, yellow, blue, red, white, lime; size 4×2: yellow, and green).

13.3.2 Artificial Industry Task

In industry, assembly pieces are usually manufactured on so-called workpiece carriers, which hold the piece in a position that facilitates assembly. Because these workpiece carriers are product-dependent, we propose creating an artificial workpiece carrier from a wooden plate and screws.

We use a 30cm×24cm wooden plate. In the middle of the plate, we drill holes to fit three types of metric screw threads[25]: M5, M2, and M8 . All holes have a distance of 6cm between them, resulting in an (X/Y) position of M5(15/6), M2 (15,12), and M8 (15,18). The carrier is depicted in Figure 13.2(b). The parts to assemble are stored in eight picking bins that are arranged in a 2×4 grid. The position of the different parts is depicted in Figure 13.2(b). We use washers and nuts for each metric thread. Additionally, we use M10 nuts and washers as distracting elements.

13.3.3 Baseline: Paper-Based Instructions

As an easy-to-reproduce baseline, we took pictures of each work step from the worker's perspective. In the Lego Duplo task, we displayed the brick to pick next in the upper left corner and highlighted the position of the brick to place

[25] ISO 68-1:1998 - `http://www.iso.org/iso/catalogue_detail.htm?csnumber=3707` (last access 5th Oct. 2016)

#steps	4	8	16	32
t_{locate_part}	1.84 (1.52)	2.22 (1.57)	2.19 (1.13)	2.29 (1.36)
t_{pick}	0.86 (0.43)	0.86 (0.39)	0.87 (0.42)	0.97 (0.63)
t_{locate_pos}	1.40 (1.07)	1.26 (1.05)	1.17 (0.48)	1.19 (0.75)
$t_{assemble\_x}$	1.30 (1.63)	0.95 (0.43)	0.96 (0.47)	0.96 (0.50)
# errors	0.17 (0.58)	0.33 (0.65)	0.33 (0.65)	0.5 (0.52)

Table 13.1: Average time in seconds for the task-independent and task-dependent measures for the Lego Duplo task and the average number of errors made. Standard deviation is depicted in parenthesis.

using a red arrow. In the industrial assembly task, we highlighted the position of the nut or washer that needs to be picked next using a red rectangle. Further we highlighted the position to assemble the part using a red circle. The pictorial instructions can be downloaded by other researchers from our website[26].

13.3.4 Evaluation

For evaluating the GATM and for providing baseline values for our suggested experiment, we conducted a user study with 12 participants (6 male, 6 female). The participants were in the age range from 22 to 31 ($M = 24.83$, $SD = 3.15$) and were recruited via a mailing list. Participants were undergraduate students with various majors. The study took approximately 30 minutes, including assembly tasks and filling out the questionnaires.

We statistically compared the average time for $t_{assemble_x}$, t_{pick}, t_{locate_part}, and t_{locate_pos}, the number of errors, and the NASA-TLX between the two proposed tasks using a one-way repeated measures ANOVA. The average times are depicted in Table 13.1 for the Lego Duplo task, and in Table 13.2 for the industrial assembly task.

The mean time to locate a part t_{locate_part} over all step-sizes was 2.14s ($SD = 0.88$s) for the Lego Duplo task and 2.17s ($SD = 0.47$s) for the industrial assembly task. A repeated measures ANOVA did not reveal a significant difference ($p > 0.05$) between the tasks. Further, we compared t_{locate_part} within each task across the different step sizes. However, the test did not find a statistically significant difference between the two tasks (all $p > 0.05$).

[26] Paper instruction download: `http://www.hcilab.org/ar-instruction-benchmark` - last access 5th Oct. 2016

#steps	4	8	16	32
t_{locate_part}	2.28 (1.09)	2.34 (1.61)	2.14 (1.46)	2.15 (1.46)
t_{pick}	1.31 (0.58)	1.48 (0.81)	1.52 (0.99)	1.60 (1.09)
t_{locate_pos}	1.15 (0.63)	1.28 (0.61)	1.29 (0.91)	1.35 (0.85)
$t_{assemble\_x}$	6.23 (7.06)	6.16 (7.03)	5.83 (9.02)	6.01 (9.61)
# errors	0.08 (0.28)	0.16 (0.39)	0.08 (0.29)	0.92 (1.24)

Table 13.2: Average time in seconds for the task-independent and task-dependent measures for the industrial task, and the average number of errors made. Standard deviation is depicted in parenthesis.

The average time to pick a part t_{pick} over each step size was 0.89s ($SD = 0.20$s) for the Lego Duplo task and 1.46s ($SD = 0.34$s) for the industrial assembly task. A repeated measures ANOVA showed a significant difference for t_{pick} between the two tasks $F(1,11) = 33.22$, $p < .001$. Further, we compared t_{pick} within each task across the different step sizes. However, a repeated measures ANOVA did not find a significant difference for both tasks (all $p > 0.05$).

The average time to find the location of a picked part t_{locate_pos} over each step-sizes was 1.25s ($SD = 0.29$s) for the Lego Duplo task and 1.25s ($SD = 0.35$s) for the industrial assembly task. A repeated measures ANOVA did not reveal a significant difference ($p > 0.05$). We further compared t_{locate_pos} within each task across the different step sizes. For the Lego Duplo task, we did not find a significant difference between the different step sizes ($p > 0.05$). Additionally, we compared the different step-sizes within the tasks using a one-way repeated measures ANOVA. Mauchly's test showed that the sphericity assumption was violated ($\chi^2(5) = 12.927$, $p = .025$). Therefore, we used the Greenhouse-Geisser correction to adjust the degrees of freedom ($\varepsilon = .62$). Interestingly, we found a significant difference between the step sizes in the industrial assembly task $F(1.887, 20.753) = 3.952$, $p = 0.037$. Pairwise comparisons revealed a significant difference between 4 and 8 steps.

The average time to assemble a part $t_{assemble_x}$ over all step sizes was 1.04s ($SD = 0.29$s) for the Lego Duplo task and 6.04s ($SD = 1.27$s) for the industrial assembly task. A repeated measures ANOVA revealed a significant difference for $t_{assemble_x}$ between the two tasks $F(1,11) = 212.32$, $p < 0.001$. We compared $t_{assemble_x}$ within each task across the different step sizes. However, a repeated measures ANOVA did not find a significant difference for both tasks (all $p > 0.05$).

We also compare the number of errors made across the different tasks. The mean number of errors made across the four different levels of complexity was 1.33 (SD

$= 1.67$) for the Lego Duplo task, and 1.25 ($SD = 1.54$) for the industrial assembly task. The analysis did not reveal a significant difference in the mean number of errors between the tasks ($p > 0.05$). Further, we compared the number of errors within each task across the different step sizes. The number of errors for each complexity is shown in Table 13.1 for the Lego Duplo task, and in Table 13.2 for the industrial assembly task. However, the ANOVA did not reveal a statistically significant difference (all $p > 0.05$).

Considering the perceived cognitive load according to the average NASA-TLX score [77] the Lego Duplo task reached a score of 33.16 ($SD = 15.52$), whereas the industrial assembly task reached a score of 43.91 ($SD = 16.77$). The analysis revealed a significant difference regarding the perceived cognitive load between the two tasks $F(1,11) = 13.05$, $p = 0.004$.

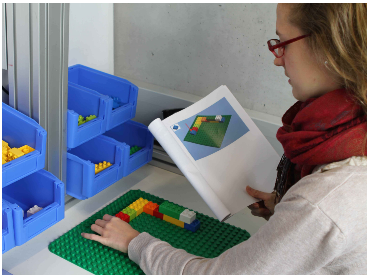
13.4 Discussion

The results of the study show that both proposed *task-dependent* measures $t_{assemble_x}$ and t_{pick} are significantly different between the two tasks. Accordingly, the two *task-independent* measures t_{locate_part} and t_{locate_pos} were not significantly different between the tasks using paper-based instructions. The results support our proposed GATM.

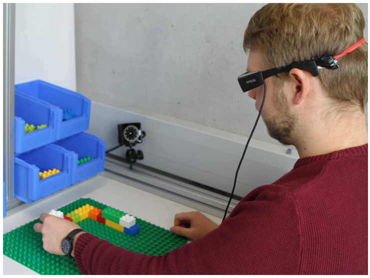
Further, we found a significant difference in t_{locate_pos} between the 4 and 8 steps industrial assembly tasks. We assume that this difference occurred because participants worked at a faster pace when assembling the 4 steps scenario. One participant (P8) stated that *"[He] wanted to finish very quickly when seeing that the instruction only consists of a little number of steps."* For all other measures and all other step-sizes the analysis did not find a significant difference. This suggests that the two assembly tasks and the used dependent variables are step-size independent and that experiments consisting of 8 work steps might be sufficient.

13.5 Evaluating Instruction Systems

After evaluating the GATM and the proposed reference tasks, we are applied the method to state-of-the-art AR instruction systems and our assistive system for a single workplace. Informed by related work, we identified four main categories of instruction systems: First, HMD-based instruction systems, where a user is

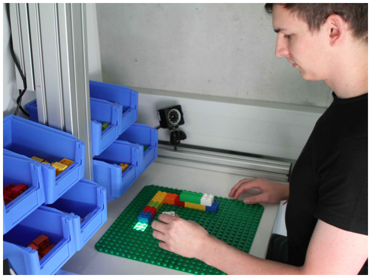


(a) Paper-based instructions



(b) HMD instructions

(c) Tablet-based instructions



(d) In-situ instructions

Figure 13.3: We used four different instruction systems in our user study. (a) A printed paper-based instruction as a baseline, (b) a digital instruction that is presented at the center of a head-mounted display, (c) a digital instruction that is presented on a tablet, and (d) in-situ projected instructions that highlight the assembly position using a green contour that is projected directly in the workspace.

viewing instructions using smart eye-wear. Second, tablet-based instructions, where a user carries a tablet containing assembly information. Third, assembly instructions using in-situ projection, where the information is directly projected onto the physical world. Lastly, many research projects still use paper-based instructions as a baseline to compare them against interactive instructions. To find the most suitable instruction system, we conducted a user study to evaluate the four instruction systems at an assembly workplace. As an assembly task for the study, we used the 32 step reference task suggested previously. In the following, we describe the instruction systems that we used in the study in detail.

Paper Instruction

As the paper baseline, we printed the reference instruction for the 32 step task that was introduced previously in this chapter on an A4 sheet of paper (see Figure 13.3(a)). We printed the instruction single-sided, that the position of the instruction is always at the same position relatively to the manual. However, this requires the worker to turn pages after each work step. Finally, we put the paper sheets together in correct order using a folder. The instruction shows a picture of the brick that needs to be picked in the upper left corner of the page. Further, the instruction shows the assembly position of the brick. To better view the assembly position, it is highlighted using a red arrow.

HMD Instruction

Inspired by Zheng et al. [165], we present pictorial instructions at the center of a HMD's field of view. To reproduce their setup, we use an Epson Moverio BT-200. The HMD displays the same images as depicted in the paper-based instruction using a full screen application. We connected the HMD via WiFi to enable a wizard of oz to advance the instruction when the assembly step was performed correctly. To ensure that the Epson Moverio BT-200 does not slip on the participant's nose, we reinforced the frame of the HMD using a rubber band (see Figure 13.3(b)).

Tablet Instruction

As a digital alternative to the paper instruction, we display assembly instructions on a tablet (see Figure 13.3(c)). Therefore, we use a HTC Nexus 9 to display images of the instruction on the tablet. We use the same instruction images that are printed in the paper-based instruction and displayed on the HMD. The tablet is connected to WiFi to enable controlling the shown instruction using a wireless presenter.

In-situ Instruction

We further use an in-situ instruction for displaying assembly information. Therefore, we use our assistive system for the single workplace that was introduced in Chapter 6 using a top-mounted projector and a top-mounted Kinect_v2. The Kinect_v2's depth data is used to detect picks from boxes and to detect if a part was assembled correctly. The projector displays a green light to highlight boxes to pick from. Accordingly, a green light is used to highlight the assembly position by projecting the contour of the part directly at the assembly position (see Figure 13.3(d)).

13.5.1 Design

We designed the experiment following a repeated measures design with one independent variable, i.e. the used system that provides assembly instructions. As dependent variables, we considered the number of errors, the perceived cognitive load using the NASA-TLX questionnaire [77], and the four different components of the TCT according to the GATM: t_{locate}, t_{pick}, t_{locate_pos}, $t_{assemble_x}$. To prevent a learning effect, we counterbalance the order of the conditions according to the Balanced Latin Square.

13.5.2 Apparatus

The workplace that was used for the study consists of two areas. First, the spare part area, which contains eight blue bins which store the parts that are used in the assembly. Second, the assembly area where the parts are assembled (see Figure 13.3), which is limited by a green Lego Duplo plate to hold the assembly in a fixed position. Previous work recognized that Lego Duplo tasks are suitable for evaluating instruction systems [14, 138, 149]. Thus, we are using the Lego Duplo reference task suggested previously, consisting of 32 steps. The task requires eight different Lego Duplo bricks that differ in color or shape. For the assembly area, we are using a green 24×24 Lego Duplo plate. In spare part area, we arrange the boxes in a 2×4 grid. We chose to use the identical 32 step task for each instruction system to ensure the same complexity for every task. To prevent a systematic learning effect, we counterbalanced the order of the instruction systems across the participants.

13.5.3 Procedure

After explaining the course of the study and signing the consent form, we collected the demographic information. To make the participants familiar with the used instructions, we gave the participants an introduction for each type of instruction directly before using it. The participants were instructed that the first priority of the study is to not make any errors, and the second priority is to assemble fast. For making the participants familiar with each type of instruction system, we used a different task than the one used in the study. When the participants felt familiar with the instruction system, the researcher started recording the assembly using a GoPro Hero3. We recorded the assembly to determine the exact times for t_{locate}, t_{pick}, t_{locate_pos}, $t_{assemble_x}$ proposed previously. In the in-situ condition and in the HMD condition the measuring of the time started with displaying the first step of the instruction. In the paper and tablet condition, the measuring started when handing the participant the instruction, as the instruction for the first step was shown right away. During the assembly, two researchers independently counted the errors that were made. We considered an error, when a wrong brick was picked and when a brick was assembled at a wrong position on the plate. In the tablet condition, HMD condition, and paper condition, we told the participants that the exact position of the bricks on the green plate is not important. They were instructed to start the assembly in the middle of the plate. Therefore, we only counted a position as wrong, if the brick was at a wrong position relatively to the other placed bricks. However, in the in-situ projection condition, we counted a wrong absolute positioning of a brick as an error, as the projection showed a fixed starting point. After the task was conducted, the researchers compared the counted number of errors. In case the number of errors differed between the researchers, they watched the recorded video and found a consent. After each condition, we asked the participants to file a NASA-TLX [77] questionnaire. Then we asked them for their opinion about the instruction system. We repeated the procedure for the other conditions. Overall, the study took approximately 30 minutes.

13.5.4 Participants

We recruited 16 participants (7 female, 9 male) via our university's mailing list. The participants were aged from 20 to 33 ($M = 25.43$, $SD = 3.59$). All participants were students with various majors or PhD students. They were not familiar with the assembled Lego Duplo task. Participants were rewarded with candies for participating in our study.

13.5.5 Results

We statistically compared the TCT divided into t_{locate}, t_{pick}, t_{locate_pos}, $t_{assemble_x}$, the number of errors, and the NASA-TLX score between the instruction systems

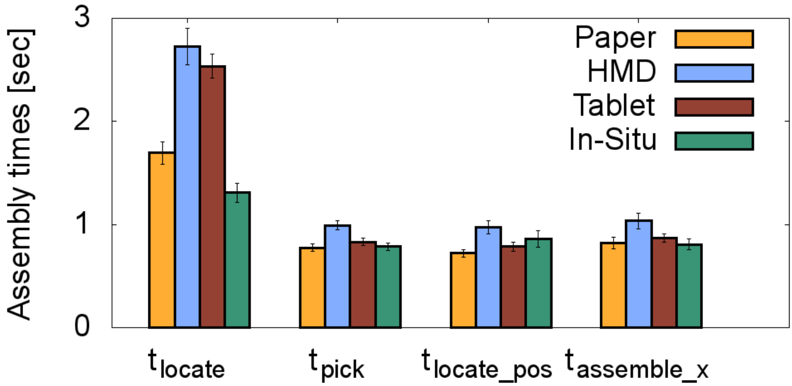
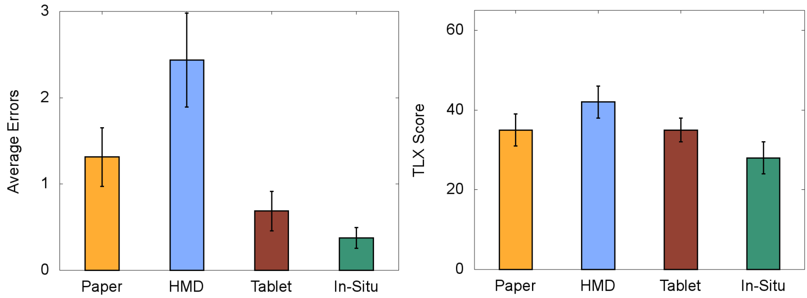


Figure 13.4: The average time for each type of instruction according to the GATM: t_{locate}, t_{pick}, t_{locate_pos}, $t_{assemble_x}$. Error bars depict the standard error.

using a one-way repeated measures ANOVA. Mauchly's test showed that the sphericity assumption was violated for the number of errors ($\chi^2(5) = 13.013$, $p = .024$), and t_{locate_pos} ($\chi^2(5) = 13.765$, $p = .017$). Therefore, we used the Greenhouse-Geisser correction to adjust the degrees of freedom ($\varepsilon = .668$ for number of errors, and $\varepsilon = .723$ for t_{locate_pos}). We further used a Bonferroni correction for all post-hoc tests.

First, we analyzed the average time that participants needed to find the bin to pick from regarding the different instructions: t_{locate}. According to Figure 13.4, the in-situ projection required the least time to process the instruction with an average of 1.30s (SD = .37s), followed by the paper instructions with an average of 1.69s (SD = .44s), the tablet instructions 2.53s (SD = .46s), and the instructions on the HMD with an average of 2.72s (SD = .72s). A one way repeated measures ANOVA revealed a significant difference between the approaches, $F(3,45) = 30.784$, $p < .001$. The post-hoc test revealed a significant difference (all $p < .05$) between all approaches except HMD instructions vs. tablet instructions and in-situ projection vs. paper instructions. The effect size estimate shows a large effect ($\eta^2 = .672$).

Considering the average time a participant needed to pick a part from the box t_{pick}, the paper instructions required the least time .77s (SD = .13s), followed by the in-situ projected instructions .78s (SD = .14s), the tablet instructions .82s (SD = .14), and the instructions that were displayed on the HMD .99s (SD = .18). The one-way repeated measures ANOVA revealed a significant difference between



(a) Number of errors (b) NASA-TLX score

Figure 13.5: (a) The average number of errors that were made during the study using the different instruction systems. (b) The average perceived cognitive load (NASA-TLX score) that was perceived by the participants when using the different instruction systems. Error bars depict the standard error.

the types of instruction, $F(3,45) = 13.362$, $p < .001$. The post-hoc test revealed that the feedback on the HMD is significantly worse than all other approaches (all $p < .05$). The effect size estimate shows a large effect ($\eta^2 = .471$).

The average time to locate the assembly position of a part t_{locate_pos} was the lowest using the paper instructions .72s ($SD = .14$s), followed by the tablet instructions .78s ($SD = .17$s), the in-situ projected instructions .85s ($SD = .32$s), and the instructions that were presented on the HMD with .97s ($SD = .26$s) on average. The one-way repeated measures ANOVA revealed a significant difference between the types of instruction, $F(2.170, 32.551) = 7.988$, $p = .001$. The post-hoc test revealed a significant difference between HMD instructions vs. paper instructions, and HMD instructions vs. tablet instructions (all $p < .05$). The effect size estimate shows a large effect ($\eta^2 = .347$).

Regarding the average time to assemble a part $t_{assemble_x}$, the in-situ projected instructions resulted in the fastest assembly with an average of .80s ($SD = .20$s), followed by the paper instructions .81s ($SD = .23$s), the tablet instructions .86s ($SD = .17$s), and the instructions that were presented on the HMD 1.03s ($SD =$ .31s). The one-way repeated measures ANOVA revealed a significant difference between the types of instruction, $F(3,45) = 6.182$, $p = .001$. The post-hoc test revealed a significant difference between the in-situ projected instructions and the instructions that are presented on the HMD. The effect size estimate shows a large effect ($\eta^2 = .292$).

Analyzing the average number of errors that were made during the assembly, the in-situ projection led to the least number of errors with .37 (SD = .50) errors on average, followed by the tablet instructions with an average of .69 (SD = .94) errors, the paper-based instructions with an average of 1.31 (SD = 1.40) errors, and the instructions on the glasses with 2.44 (SD = 2.25) errors on average. The one-way repeated measures ANOVA showed a significant difference between the approaches, $F(2.005, 30.070) = 7.859$, $p = .002$. The post-hoc test revealed a significant difference between the HMD instructions vs. the tablet instructions, and the HMD instructions vs. the in-situ projection. The effect size estimate shows a large effect ($\eta^2 = .344$). The results are also depicted in Figure 13.5(a).

Considering the perceived cognitive load using the NASA-TLX score [77], the in-situ projection was perceived best with an average TLX score of 28.13 (SD = 18.41), followed by the tablet with an average of 35.06 (SD = 15.48), the paper baseline with an average of 35.50 (SD = 18.19) and the HMD with an average score of 42.81 (SD = 18.28). A one-way repeated measures ANOVA revealed a significant difference between the approaches, $F(3,45) = 5.171$, $p = .004$. The post-hoc test only revealed a significant difference ($p < .05$) between the HMD and the in-situ instructions. The effect size estimate revealed a medium effect ($\eta^2 = .256$). A graphical representation is depicted in Figure 13.5(b).

Qualitative Results

We analyzed the statements of the participants after assembling the Lego Duplo bricks according to each of the instruction systems. For the HMD instructions, the participants mostly stated that *"the displayed instruction blocks the sight on the assembly and the bins that contain the bricks"* (P2, P3, P4, P7, P9, P11, P12). Further, they told us that *"when I am focusing a point that is very close, the HMD shows two pictures, which makes the instruction hard to see"*. However, they liked that *"the HMD instructions enabled [them] to assemble with both hands"*. Considering the tablet instructions, participants liked that *"compared to the paper instruction, there is no chance that [I] skip a page unintentionally"* (P13). In contrast, P4 stated that *"the tablet interferes with the assembly task as I can only use one hand"*. The paper instructions were perceived well by the participants, as they *"can put [them] away if [they] don't need them anymore"* (P14, P16). However, a participant stated that he *"needed to double check if [he] didn't skip a page"* (P13). Finally, for the in-situ instructions, the participants liked that they *"have both hands free during the assembly"* (P4, P8) and that they *"don't need to think to transfer the instruction to the work area"* (P2, P12, P14). In contrast, they remarked that *"the projection could be brighter, as it was hard to notice on blue bricks"* (P7).

13.6 Summary

In this chapter, we introduce the General Assembly Task Model using *task-dependent* and *task-independent* measures and provide a uniform experiment design as a benchmark for evaluating assembly instructions. Further, we suggest two reference tasks that are cheap to build, easily reproducible, and that are covering most tasks that are found in assembly workplaces in industrial settings. We also provide a paper-based baseline for the two proposed tasks that can be downloaded by other researchers. We are confident that introducing this uniform benchmark will make different instruction approaches more comparable to each other.

Further, we used one of the proposed reference tasks and the GATM method to evaluate different instruction systems for providing assembly instructions at the workplace. We compared centrally-positioned HMD instructions, to tablet instructions, in-situ projected instructions, and paper instructions. Considering the assembly times, our results show that locating a part is significantly faster using in-situ projection and paper-based instructions, picking a part is significantly slower using central HMD instructions compared to other instructions, locating assembly positions is significantly slower using HMD instructions compared to tablet and paper instructions, and assembling is significantly faster using in-situ projection compared to instructions that are shown on an HMD. Further, participants made significantly less errors using the tablet and in-situ instructions compared to the HMD instructions. Moreover, the perceived cognitive load using the NASA-TLX [77] questionnaire is significantly lower for the in-situ instructions compared to the HMD instructions. Participants liked that they have their hands free when using in-situ instructions. However, they found that HMD instructions block their field of view and holding tablet or paper instructions during assembly tasks interferes with assembling using both hands.

Although the paper baseline was not significantly worse than in-situ projection, we assume that the hands-free character of in-situ projection has a greater potential for instruction systems at the workplace. This is also because workers find HMDs difficult to work with and tablet instructions interfere with a two-handed assembly.

V

Results

Chapter 14

Guidelines for Assistive Systems

This chapter is based on the following publication:

- M. Funk, T. Kosch, R. Kettner, O. Korn, and A. Schmidt. motioneap: An overview of 4 years of combining industrial assembly with augmented reality for industry 4.0. In *Proceedings of the 16th International Conference on Knowledge Technologies and Data-driven Business*, i-KNOW '16, New York, NY, USA, 2016. ACM

Based on the experience we gained from using our assistive system in different scenarios and with different user groups, we propose guidelines for designing and implementing assistive systems for the workplace. Although these guidelines and recommendations were inspired by using assistive systems at workplace scenarios, we suggest that they can be transferred to other scenarios involving assistive systems, e.g. for a smart kitchen scenario (cf. [52]).

1. Keep the feedback simple:
When designing instructions for assistive systems, the designer should consider two important aspects: First, the instructions should contain all important information that is relevant for the task. Second, the instructions should be as simple as

possible. As the two aspects are contrary, the minimum trade-off that still fulfills both requirements needs to be found. In our studies, we showed that displaying text should be avoided as some workers are not able to read. Also there might be foreign workers, who are not able to understand a language. Further, we found that videos and complex pictures should also be avoided as they transfer a lot more information than what is necessary to complete the task. As the best trade-off between the two requirements, we found that displaying contour information for showing assembly steps is a good way of fulfilling both requirements (see Chapter 8).

2. Display direct feedback:
Assistive systems have many possibilities to display feedback and instructions to users. Through our experience with assistive systems, we learned that feedback should be displayed directly at the position where the action is required. When displaying feedback on a screen that is located in close proximity to where the actions is required, users have to transfer the instruction that is shown on a screen to the real world. This requires cognitive effort, consumes time, and is error-prone. In our studies, we found that using in-situ projection to display projected feedback directly on the area where the action is required, is faster and less cognitively demanding.

3. Design for context-aware usage:
Operating an assistive system should not result in additional effort and should not limit the user in performing their tasks. Therefore, we require assistive systems to be designed to enable a hands-free usage. If the user would have to carry a remote controller for interacting with the assistive system, the user's hands would always be occupied by using the controller. Therefore, we argue for using activity recognition for interacting with assistive systems while performing tasks. Based on the recognized actions, instructions and feedback can be displayed. While in a paper-based reference manual the correct page that matches with the action to perform needs to be found first, an interactive system can select the appropriate feedback or instruction based on context. If a configuration action aside from the regular work task needs to be triggered, touch screens or gestures can be used.

4. Equip the environment rather than the user:
When analyzing the design space of assistive systems, the dimension of where to put the technology is important. Assistive systems can be stationary and mounted in the environment or can be mobile and carried by the user. Also hybrid approaches exist, where assistive systems are both placed away from the user, but the user can carry the technology on demand (e.g. the OrderPickAR cart introduced in Chapter 7). Through many studies with experienced workers and

cognitively impaired workers, we recommend to rather equip the environment with technology of the assistive system than equipping users with technology. We learned that users do not want to wear any additional piece of technology when performing a work task. Further, the users leave the work area to perform other tasks. Then the users would have to take off the technology and put it on again when resuming the task. Also when assistive systems are placed in the environment, multiple users can benefit from it without exchanging technology between the users.

5. Strive for intuitive natural interaction:
For integrating an assistive system seamlessly into a scenario in the physical world, all interaction with the assistive system also should happen in the physical world, and not on a configuration device. We distinguish two scenarios: interacting with the system regularly and programming procedures or workflows for the system. As users of assistive systems sometimes are not able to use a computer or cannot deal with a complex GUI, interaction with the system should happen based on natural interaction and detecting activity. Further, as designers of assistive systems, we should keep in mind that also users who are teaching workflows or procedures to assistive systems usually do not have a programming background. Although these users are experts in the task that they are teaching, they are not necessarily experts in using computer systems. Therefore, we recommend to also use natural interaction for programming assistive systems. One concept that is suggested in this thesis for using natural interaction for programming workflows is the Programming by Demonstration concept (see Section 4.4).

6. Design for personalized feedback:
During our studies, we explored the best feedback for both communicating work instructions and for presenting errors. Although we found that a contour visualization of assembly positions is the best way to communicate assembly instructions, there were still users that preferred watching an assembly video or looking at pictures of the assembly. Also for communicating errors that happen during the assembly process, we found that a combination of visual and haptic feedback is a superior trade-off between privacy and error-awareness compared to the other modalities. Despite these results, some users of the assistive system preferred auditory feedback over the visual and haptic feedback. In contrast, cognitively impaired users liked it a lot when they received positive feedback after a work step was performed. As designers of assistive systems, we should take these aspects into account. For assistive systems at the workplace, the standard feedback that is presented to the average worker should represent the best feedback for the overall population. However, we should provide the opportunity

to personalize the feedback that is given and adjust it to the preferences of the users.

7. Enable users to control the speed:
Particularly when using assistive systems for augmenting work processes, it is important not to rush the user. Instead, the users should be able to perform the steps according to their own pace. This results in a important design decision: all actions that advance instructions or feedback should be triggered by the user and not by the system. Only if the user initiates the process, the user feels in control (cf. [145]). If the system would proceed instructions after a defined amount of time, the system would set the pace of the task. We assume that this would lead to more stress at the workplace. Additionally, for explicitly interacting with the system, e.g. for skipping a work step or for replaying the previous work step, we integrated a foot pedal that can be pressed by the user to control the feedback that is currently displayed.

8. Add motivating quantified-self information:
During our studies, participants occasionally asked how many items were left in the current task and how fast they were. Some participants suggested to have this information always present while performing their work tasks. This could be as simple as displaying a progress bar which fills up when completing more work steps, or more complex with a leader-board that shows which worker made the fewest errors or which worker produced the most parts. Displaying quantified-self information can be closely linked to adding gamification elements to enhance work processes, which was suggested by Korn et al. [95, 96, 99]. We are convinced that designing assistive systems in a way that users can always view their quantified-self information will lead to a higher motivation during monotonous tasks when using assistive systems.

Chapter 15

Conclusion and Future Work

This thesis explores how Augmented Reality instructions can be used to support inexperienced workers, experienced workers, and workers with cognitive impairments during assembly or order picking tasks. To investigate this, we followed a user-centered design process with multiple iterations and followed a bottom-up approach. Thereby, we first collected the workers' requirements and then built a system to understand the workers' needs in more depth. In this chapter, we summarize the research contributions by answering the research questions that were addressed in this thesis. Further, we point towards next steps for future work.

15.1 Summary of Research Contribution

The main contribution of this thesis is three-fold. First, we applied a bottom-up user-centered design process with multiple iterations for identifying requirements and building an assistive system for the workplace that can be applied to both manual assembly workplaces and order picking scenarios. Second, we present four prototypes implementing the identified requirements and exploring how to apply the prototypes in different design dimensions of the design space. For the assembly workplace, we present a single assembly workplace and an assembly cell connecting three workplaces. Considering the order picking scenario, we present a user-mounted and a cart-mounted system. Third, we formulate and present guidelines for designing assistive systems and for designing the presented instructions for workplaces. We base these guidelines on our research and experience with assistive systems for the workplace.

15.1.1 Requirements for Assistive Systems

By applying the user-centered design process, we first identified use-cases for assistive systems at the workplace and found two major scenarios: training workers for a new task and providing continuous quality support for workers with cognitive deficits. For these scenarios, we identified eight functional requirements and four non-functional requirements. As we chose a bottom-up approach for implementing the prototypes, we first defined a general architecture, which fulfills each requirement, and afterwards built the prototypes. Within the constraints of the requirements, we further present three interaction concepts that are explored in the context of this thesis: Programming by Demonstration (PbD), implicit interaction with assistive systems, and using user-defined tangible objects.

15.1.2 Assistive System Prototypes

One main contribution of this thesis is the developed software architecture and the resulting prototypes for assistive systems supporting both manual assembly and order picking. As the software architecture was designed for being generally applicable to a variety of hardware setups, we demonstrated this flexibility by applying our software to four research prototypes. First, we described the generic parts of the software architecture by introducing a concept for workflows and visual feedback using *Scenes and Triggers*. Further, we designed a concept for making the instructions adaptive, based on a worker's skill level. For enabling implicit interaction with the assistive system, we introduce three activity recognition techniques: pick detection, assembly detection, and object recognition. For providing feedback, we introduce a concept for communicating visual instructions using in-situ projection. Finally, we apply the introduced architecture and the instruction concepts to four different assistive systems: a single workplace for supporting manual assembly tasks, a u-shaped assembly cell simultaneously supporting three workers during manual assembly tasks, an augmented picking cart which provides in-situ picking instructions for order picking tasks, and an augmented picking helmet which provides user-mounted order picking support. By applying the architecture to these four different systems, we show the flexibility and general applicability of our software.

15.1.3 Guidelines for Designing Assistive Systems

As a result of designing, implementing, and evaluating assistive systems for approximately four years, we presented eight guidelines for designing assistive systems. They should help designers and researchers to understand the concepts that worked well with the users and the concepts that did not work well with the users. We based these guidelines on our experience that we gathered from applying our assistive system in many different industrial scenarios e.g. in an assembly hall of a major car manufacturing company, or in a sheltered work organization for supporting workers with cognitive impairments. Additionally, by exhibiting our assistive system at several trading fairs, we discussed with a number of experts for manual assembly and experts for assistive systems, which helped us to further shape the guidelines.

15.2 Research Questions

At the beginning of this thesis, we stated seven research questions that were addressed in the evaluation chapters. In this section, we answer the research questions based on the knowledge we gained from conducting user studies and interviews.

RQ1: What are suitable in-situ visualizations for assembly instructions?

As an assistive system using in-situ projection can display nearly any visual content at the workplace, the first and most important research question was which visualizations of in-situ instructions are suitable for communicating assembly instructions. Through a user study with 64 cognitively impaired workers, we compared three common visualizations, namely, a contour visualization, a pictorial visualization, and a video visualization, to a baseline without any instruction (cf. Chapter 8). We discovered that using a contour visualization led to the shortest assembly times and the least errors. Further, the contour visualization led to a significantly lower perceived cognitive load compared to not using any visual feedback. Additionally, the contour visualization performs equally well for all degrees of cognitive impairments. A contour visualization can communicate the position and orientation of the part that needs to be assembled, without introducing complex graphics or visual clutter. Therefore, we argue that a contour visualization is the most suitable visualization for communicating in-situ instructions.

RQ2: How are in-situ instructions perceived by cognitively impaired workers?

We learned that in-situ instructions can have a positive or a negative effect depending on the visualization. We observed that cognitively impaired workers took a longer time to perform manual assembly tasks using an in-situ video instruction than using no visual instruction, however the difference was not statistically significant. In contrast, contour visualizations were able to improve the assembly time. We further investigated and quantified, to which extent the in-situ projected contour instructions change the assembly time and the errors by conducting a study, where we compared contour instructions to state-of-the-art pictorial instructions with different assembly complexities (cf. Chapter 8). Considering the time to assemble, we found that contour-based in-situ instructions are significantly faster in all assembly complexities from 6 work steps up to 96 work steps. When analyzing the errors that were made during the assembly, we found a significant effect between the instructions starting from 24 work steps. This is an important finding, considering that workers with cognitive impairments usually only work on rather simple assembly tasks with 5.25 work steps on average. Thus, we are confident that contour-based in-situ instructions have the potential to change the way sheltered work organizations are organizing their assembly tasks and can enable workers with cognitive impairments to work on more complex assembly tasks.

RQ3: Which modality can communicate errors best to workers?

As assistive systems are able to detect when a worker is making an error, another important research question was which modality to use for communicating to the workers that an error was made. Through a user study with non-impaired workers, we compared haptic, auditory, and visual error feedback (cf. Chapter 9). We first identified the most suitable variant for each modality in a pre-study. In a main study, we then compared the best variants and their combinations. By analyzing an overall subjective rating, the perceived privacy violation, and the perceived distraction, we found that a combination of visual and haptic feedback was most suitable for workers without impairments. Beyond the scope of this thesis, we found that for workers with cognitive impairments only visual error feedback is suitable [101].

RQ4: How can in-situ instructions for assistive systems be created?

When each workplace in manual assembly is equipped with an assistive system offering in-situ instructions, creating these instructions is a crucial part. As we are interested in how these instructions can be created easily, we compared existing instruction creation technologies (recording an assembly video or using a graphical editor) to our proposed Programming by Demonstration (PbD) instruction creation method (cf. Chapter 10). When analyzing how to create instructions, two aspects are crucial: how are the created instructions perceived by the workers, and how comfortable is the creation of the instructions for the trainer. By evaluating the performance of the users, we found that in-situ instructions outperform video-based instructions in task completion time, number of errors, and NASA-TLX. Regarding the creation of the instructions, we found that creating instructions using PbD takes approximately 3 times longer than recording a video of the assembly. On a meta level, the in-situ instructions are better for the workers, as the created instructions contain depth information that can be used for checking if parts are correctly assembled. Additionally they lead to a faster assembly compared to video-based instructions. Considering the time a worker takes for assembling a product, taking three times longer for creating an in-situ instruction might the better solution with the whole assembly process in mind.

RQ5: What are the long-term effects of using in-situ instructions?

In an ideal production scenario, the products are manufactured on demand in lot size one. As producing in lot size one results in a high cognitive effort for the workers, it is of interest to explore the long-term effects of assistive systems providing cognitive support. In a user study with experienced and inexperienced workers, we explored the effects of using in-situ instructions for three work days in a row (cf. Chapter 11). We found that using in-situ instructions for a longer period of time slows down the workers. After the task was learned by the workers, the instructions were perceived as visual distraction, which lead to a longer assembly time. This emphasizes the need for adaptive instructions that can adjust to the workers' cognitive needs.

RQ6: How can in-situ instructions be used for order picking tasks?

As a different scenario, we investigated order picking tasks for providing cognitive support using in-situ instructions. For transferring the concepts to order picking, we implemented two prototypes OrderPickAR and HelmetPickAR, which investigated placing the equipment on an order picking cart and equipping the user (cf. Chapter 12). For the OrderPickAR system, we found that using in-situ projection for order picking tasks has a great potential, as the in-situ instructions could significantly outperform the paper baseline considering task completion time and perceived cognitive load. We also learned that the control condition using an HMD for providing order picking instructions was not accepted by the workers. With the HelmetPickAR prototype, we explored the user-mounted design dimension and found that HelmetPickAR lead to significantly lower perceived cognitive workload compared to a paper baseline. Overall the two prototypes introduced in this thesis could show that there is potential for using in-situ projection in order picking tasks. As the feedback is directly projected onto the environment, the cognitive effort for transferring the instruction to the environment is reduced.

RQ7: How can instructions for workplaces be evaluated?

During the course of this thesis, we found that there is a need for evaluating Augmented Reality instructions in a standardized way. As each research project introduces its own task for each conducted study, the Augmented Reality instructions that can be found in the literature are barely comparable to each other. This is mainly because the papers report the task completion time as a dependent measure, which is influenced by the used assembly task. To provide a standardized way of evaluating Augmented Reality instructions, we introduce two reference tasks and a way of evaluating the instructions to reduce task-dependent noise. We called this technique the General Assembly Task Model. Through a user study, we compared in-situ instructions to a paper baseline, instructions shown on an HMD, and instructions shown on a tablet computer (cf. Chapter 13). In the study, we found that in-situ projection is the fastest and least error-prone instruction that is available, since no transformation has to be done by the user to map the information in the instructions onto the physical work place.

15.3 Future Work

In this thesis, we investigated the use of in-situ projection for providing cognitive assistance in workplace scenarios. As the conducted research and the presented prototypes build a solid foundation for the workplace scenario, we identified further interesting areas of research that were beyond the scope of this thesis. In this section, we present these areas for future research.

15.3.1 Exploring the Memorability of In-Situ Instructions

As we found that in-situ projection can be used to train inexperienced workers for learning new assembly tasks, it would be interesting to quantify to which extent in-situ projection can be used for learning. A research project could compare the memorability of in-situ instructions to the memorability of paper-based instructions. This could be done by conducting a lab study, where participants have to learn how to assemble a product using both instruction types. As a further step, the memorability of instructions could be explored in real manufacturing scenarios, e.g. using both instruction types to train new workers in car manufacturing and quantify their learning progress.

15.3.2 Extending In-Situ Projection Interfaces to Other Application Areas

We identified that there is a huge potential for supporting workers with cognitive impairments at workplaces. In future work, this concept could be extended to support persons with cognitive impairments in other areas of living. One example could be using in-situ projection to prepare meals in the kitchen or to use in-situ projection for helping persons with cognitive impairments to dress in the morning. Additionally, this concept could also be evaluated for persons without cognitive impairments. Considering a smart kitchen scenario, e.g. a grandmother could be using the Programming by Demonstration approach to create a baking instruction for the grandchild's favorite pie. Then the grandchild could use the in-situ baking instruction to bake the pie exactly like the grandmother does.

15.3.3 Towards Smart Lights: Using Interactive Ubiquitous Projection in Everyday Scenarios

In a larger context, combining sensing technology and a light source for creating interactive experiences in everyday scenarios is an interesting research area to investigate further. Light is a ubiquitous technology that can be found everywhere, e.g. on the streets, at home, at the workplace, or even as a portable flashlight. A research project could investigate how sensing technology and output technology can be combined in a light bulb. Then input and output technology would be combined in one device that can easily be deployed. Thereby, the research challenges are in identifying input and output concepts, creating a way for passersby to interact with publicly installed smart lights, e.g. to spontaneously create personal displays, and to create a cheap and small system that can be ubiquitously deployed.

15.4 Concluding Remarks

This thesis investigates how in-situ projected instructions can be used for cognitively supporting workers with and without cognitive impairments in manual assembly workplaces. Further, we investigated applying these concepts to order picking scenarios. We are confident that this technology will have a great impact on the design and planning of factories where assembly tasks are performed. Having an assistive system mounted on every assembly workplace could change the way how workers are employed and trained, and even how products are being produced. At the moment of finishing this thesis, we already see commercial products offering in-situ projection for workplace scenarios (cf. Section 2.5). With assistive systems using in-situ projection being commercially available and the promising results that this thesis provides, we assume that in some years this technology will truly become ubiquitous. To further promote research in using in-situ projection, we will provide the software that was created within this thesis as an open-source project to interested researchers and designers.

VI

Bibliography

BIBLIOGRAPHY

[1] Y. Abdelrahman, M. Hassib, M. G. Marquez, M. Funk, and A. Schmidt. Implicit engagement detection for interactive museums using brain-computer interfaces. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, pages 838–845. ACM, 2015.

[2] M. Aehnelt and B. Urban. Follow-me: smartwatch assistance on the shop floor. In *HCI in Business*, pages 279–287. Springer, 2014.

[3] M. Agrawala, D. Phan, J. Heiser, J. Haymaker, J. Klingner, P. Hanrahan, and B. Tversky. Designing effective step-by-step assembly instructions. In *ACM Transactions on Graphics (TOG)*, volume 22, pages 828–837. ACM, 2003.

[4] M. Akamatsu, I. S. MacKenzie, and T. Hasbroucq. A comparison of tactile, auditory, and visual feedback in a pointing task using a mouse-type device. *Ergonomics*, 38(4):816–827, 1995.

[5] J. Aleotti and S. Caselli. Robust trajectory learning and approximation for robot programming by demonstration. *Robotics and Autonomous Systems*, 54(5):409–413, 2006.

[6] R. Alm, M. Aehnelt, and B. Urban. Plant@ hand: from activity recognition to situation-based annotation management at mobile assembly workplaces. In *Proceedings of the 2nd international Workshop on Sensor-based Activity Recognition and Interaction*, page 15. ACM, 2015.

[7] S. Antifakos, F. Michahelles, and B. Schiele. Proactive instructions for furniture assembly. In *UbiComp 2002: Ubiquitous Computing*, pages 351–360. Springer, 2002.

[8] M. Avila, M. Funk, and N. Henze. Dronenavigator: Using drones for navigating visually impaired persons. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*, pages 327–328. ACM, 2015.

[9] A. Bächler, L. Bächler, S. Autenrieth, P. Kurtz, T. Hoerz, T. Heidenreich, and G. Kruell. A comparative study of an assistance system for manual order picking–called pick-by-projection–with the guiding systems pick-by-paper, pick-by-light and pick-by-display. In *2016 49th Hawaii International Conference on System Sciences (HICSS)*, pages 523–531. IEEE, 2016.

[10] A. Bächler, L. Bächler, P. Kurtz, G. Kruell, T. Heidenreich, and T. Hoerz. A study about the comprehensibility of pictograms for order picking processes with disabled people and people with altered performance. In *Intelligent Interactive Multimedia Systems and Services*, pages 69–80. Springer, 2015.

[11] A. Bächler, P. Kurtz, T. Hoerz, G. Kruell, L. Bächler, and S. Autenrieth. About the development of an interactive assistance system for impaired employees in manual order picking. In *Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, page 4. ACM, 2015.

[12] L. Bächler, A. Bächler, M. Funk, S. Autenrieth, G. Kruell, T. Hoerz, and T. Heidenreich. The use and impact of an assistance system for supporting participation in employment for individuals with cognitive disabilities. In *International Conference on Computers Helping People with Special Needs*, pages 329–332. Springer, 2016.

[13] A. Bannat, J. Gast, G. Rigoll, and F. Wallhoff. Event analysis and interpretation of human activity for augmented reality-based assistant systems. In *Intelligent Computer Communication and Processing, 2008. ICCP 2008. 4th International Conference on*, pages 1–8. IEEE, 2008.

[14] A. Bannat, F. Wallhoff, G. Rigoll, F. Friesdorf, H. Bubb, S. Stork, H. Müller, A. Schubö, M. Wiesbeck, and M. Zäh. Towards optimal worker assistance: a framework for adaptive selection and presentation of assembly instructions. In *Proceedings of the 1st international workshop on cognition for technical systems, Cotesys*, 2008.

[15] M. Barrett, J. Blackledge, and E. Coyle. Using virtual reality to enhance electrical safety and design in the built environment. 2011.

[16] P. Baudisch and R. Rosenholtz. Halo: a technique for visualizing off-screen objects. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 481–488. ACM, 2003.

[17] H. Bay, T. Tuytelaars, and L. Van Gool. Surf: Speeded up robust features. In *Computer vision–ECCV 2006*, pages 404–417. Springer, 2006.

[18] P. Beardsley, J. Van Baar, R. Raskar, and C. Forlines. Interaction using a handheld projector. *Computer Graphics and Applications*, 25(1):39–43, 2005.

[19] H. Behrendt, M. Funk, and O. Korn. Ethical implications regarding assistive technology at workplaces. In C. Misselhorn, editor, *Collective Agency and Cooperation in Natural and Artificial Systems*, volume 122 of *Philosophical Studies Series*. Springer International Publishing.

[20] D. Bial, D. Kern, F. Alt, and A. Schmidt. Enhancing outdoor navigation systems through vibrotactile feedback. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, pages 1273–1278. ACM, 2011.

[21] A. Billard, S. Calinon, R. Dillmann, and S. Schaal. Robot programming by demonstration. In *Springer handbook of robotics*, pages 1371–1394. Springer, 2008.

[22] M. Billinghurst, M. Hakkarainen, and C. Woodward. Augmented assembly using a mobile phone. In *Proceedings of the 7th International Conference on Mobile and Ubiquitous Multimedia*, pages 84–87. ACM, 2008.

[23] F. Biocca, A. Tang, C. Owen, and F. Xiao. Attention funnel: omnidirectional 3d cursor for mobile augmented reality platforms. In *Proceedings of the SIGCHI conference on Human Factors in computing systems*, pages 1115–1122. ACM, 2006.

[24] U. Blanke, B. Schiele, M. Kreil, P. Lukowicz, B. Sick, and T. Gruber. All for one or one for all? combining heterogeneous features for activity spotting. In *Pervasive Computing and Communications Workshops (PERCOM Workshops), 2010 8th IEEE International Conference on*, pages 18–24. IEEE, 2010.

[25] R. Bokranz and K. Landau. Handbuch industrial engineering. *Schäffer-Poeschel, Stuttgart*, 2012.

[26] S. Bottecchia, J.-M. Cieutat, and J.-P. Jessel. Tac: augmented reality system for collaborative tele-assistance in the field of maintenance through internet.

In *Proceedings of the 1st Augmented Human International Conference*, page 14. ACM, 2010.

[27] A. Boud, D. J. Haniff, C. Baber, and S. Steiner. Virtual reality and augmented reality as a training tool for assembly tasks. In *Information Visualization, 1999. Proceedings. 1999 IEEE International Conference on*, pages 32–36. IEEE, 1999.

[28] S. Bundesamt. Bevölkerung deutschlands bis 2060–13. koordinierte bevölkerungsvorausberechnung. *Statistisches Bundesamt: Wiesbaden*, 2015.

[29] S. Büttner, M. Funk, O. Sand, and C. Röcker. Using head-mounted displays and in-situ projection for assistive systems - a comparison. In *Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, New York, NY, USA, 2016. ACM.

[30] S. Büttner, O. Sand, and C. Röcker. Extending the design space in industrial manufacturing through mobile projection. In *Proceedings of the 17th International Conference on Human-Computer Interaction with Mobile Devices and Services Adjunct*, pages 1130–1133. ACM, 2015.

[31] A. Butz, M. Schneider, and M. Spassova. Searchlight–a lightweight search function for pervasive environments. In *Pervasive Computing*, pages 351–356. Springer, 2004.

[32] T. P. Caudell and D. W. Mizell. Augmented reality: An application of heads-up display technology to manual manufacturing processes. In *System Sciences, 1992. Proceedings of the Twenty-Fifth Hawaii International Conference on*, volume 2, pages 659–669. IEEE, 1992.

[33] T. Collett and B. A. MacDonald. Developer oriented visualisation of a robot program. In *Proceedings of the 1st ACM SIGCHI/SIGART conference on Human-robot interaction*, pages 49–56. ACM, 2006.

[34] D. Crasto, A. Kale, and C. Jaynes. The smart bookshelf: A study of camera projector scene augmentation of an everyday environment. In *Application of Computer Vision, 2005. WACV/MOTIONS'05 Volume 1. Seventh IEEE Workshops on*, volume 1, pages 218–225. IEEE, 2005.

[35] A. J. Cuvo, P. K. Davis, M. F. O'Reilly, B. M. Mooney, and R. Crowley. Promoting stimulus control with textual prompts and performance feedback for persons with mild disabilities. *Journal of Applied Behavior Analysis*, 25(2):477–489, 1992.

[36] F. De Crescenzio, M. Fantini, F. Persiani, L. Di Stefano, P. Azzari, and S. Salti. Augmented reality for aircraft maintenance training and operations support. *Computer Graphics and Applications, IEEE*, 31(1):96–101, 2011.

[37] R. De Koster, T. Le-Duc, and K. J. Roodbergen. Design and control of warehouse order picking: A literature review. *European Journal of Operational Research*, 182(2):481–501, 2007.

[38] T. Dingler, M. Funk, and F. Alt. Interaction proxemics: Combining physical spaces for seamless gesture interaction. In *Proceedings of the 4th International Symposium on Pervasive Displays*, pages 107–114. ACM, 2015.

[39] F. Echtler, F. Sturm, K. Kindermann, G. Klinker, J. Stilla, J. Trilk, and H. Najafi. The intelligent welding gun: Augmented reality for experimental vehicle construction. In *Virtual and augmented reality applications in manufacturing*, pages 333–360. Springer, 2004.

[40] D. D. Eisenstein. Analysis and optimal design of discrete order picking technologies along a line. *Naval Research Logistics (NRL)*, 55(4):350–362, 2008.

[41] W. Erickson, C. Lee, and S. von Schrader. Disability statistics from the 2013 american community survey (acs). Retrieved Sept. 25th 2015 from `www.disabilitystatistics.org`, 2015.

[42] M. Fiorentino, R. de Amicis, G. Monno, and A. Stork. Spacedesign: A mixed reality workspace for aesthetic industrial design. In *Proceedings of the 1st International Symposium on Mixed and Augmented Reality*, page 86. IEEE Computer Society, 2002.

[43] P. Fite-Georgel. Is there a reality in industrial augmented reality? In *Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium on*, pages 201–210. IEEE, 2011.

[44] H.-J. Franke et al. Variantenvielfalt und resultierende komplexität, ursachen und methoden zu ihrer bewältigung. In *DFX 1998: Proceedings of the 9th Symposium on Design for Manufacturing, Schnaittach/Erlangen, Germany, 15.-16.10. 1998*, 1998.

[45] M. Funk, A. Bächler, L. Bächler, O. Korn, C. Krieger, T. Heidenreich, and A. Schmidt. Comparing projected in-situ feedback at the manual assembly workplace with impaired workers. In *Proceedings of the 8th International*

Conference on PErvasive Technologies Related to Assistive Environments, New York, NY, USA, 2015. ACM.

[46] M. Funk, R. Boldt, M. Eisele, T. Yalcin, N. Henze, and A. Schmidt. Visualizing locations for a search engine for the physical world. In *Mensch & Computer*, pages 205–214, 2014.

[47] M. Funk, R. Boldt, B. Pfleging, M. Pfeiffer, N. Henze, and A. Schmidt. Representing indoor location of objects on wearable computers with head-mounted displays. In *Proceedings of the 5th Augmented Human International Conference*, page 18. ACM, 2014.

[48] M. Funk, T. Dingler, J. Cooper, and A. Schmidt. Stop helping me-i'm bored!: why assembly assistance needs to be adaptive. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, pages 1269–1273. ACM, 2015.

[49] M. Funk, J. Heusler, E. Akcay, K. Weiland, and A. Schmidt. Haptic, auditory, or visual? towards optimal error feedback at manual assembly workplaces. In *Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, New York, NY, USA, 2016. ACM.

[50] M. Funk, O. Korn, and A. Schmidt. Assisitive augmentation at the manual assembly workplace using in-situ projection. In *Proceeding of the chi workshop on assistive augmentation*, page 11, 2014.

[51] M. Funk, O. Korn, and A. Schmidt. An augmented workplace for enabling user-defined tangibles. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, pages 1285–1290. ACM, 2014.

[52] M. Funk, O. Korn, and A. Schmidt. Enabling end users to program for smart environments. 12:9–14, 2015.

[53] M. Funk, T. Kosch, S. W. Greenwald, and A. Schmidt. A benchmark for interactive augmented reality instructions for assembly tasks. In *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia*. ACM, 2015.

[54] M. Funk, T. Kosch, R. Kettner, O. Korn, and A. Schmidt. motioneap: An overview of 4 years of combining industrial assembly with augmented reality for industry 4.0. In *Proceedings of the 16th International Conference*

on Knowledge Technologies and Data-driven Business, i-KNOW '16, New York, NY, USA, 2016. ACM.

[55] M. Funk, T. Kosch, and A. Schmidt. Interactive worker assistance: Comparing the effects of head-mounted displays, in-situ projection, tablet, and paper instructions. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, 2016.

[56] M. Funk, T. Kosch, K. Wolf, P. Knierim, S. Mayer, and A. Schmidt. Automatic projection positioning based on surface suitability. In *Proceedings of the 5th ACM International Symposium on Pervasive Displays*, PerDis'16, New York, NY, USA, 2016. ACM.

[57] M. Funk, L. Lischke, S. Mayer, A. Sahami Shirazi, and A. Schmidt. Assistive augmentation at the workplace: A system for creating semantically-rich assembly instructions. In *Assistive Augmentation: Cognitive Science and Technology (to appear)*. Springer, 2017.

[58] M. Funk, S. Mayer, M. Nistor, and A. Schmidt. Mobile in-situ pick-by-vision: Order picking support using a projector helmet. In *Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, New York, NY, USA, 2016. ACM.

[59] M. Funk, S. Mayer, and A. Schmidt. Using in-situ projection to support cognitively impaired workers at the workplace. In *Proceedings of the 17th international ACM SIGACCESS conference on Computers & accessibility*. ACM, 2015.

[60] M. Funk, A. Sahami, N. Henze, and A. Schmidt. Using a touch-sensitive wristband for text entry on smart watches. In *CHI'14 Extended Abstracts on Human Factors in Computing Systems*, pages 2305–2310. ACM, 2014.

[61] M. Funk and A. Schmidt. Cognitive assistance in the workplace. *Pervasive Computing, IEEE*, 14(3):53–55, July 2015.

[62] M. Funk, A. Schmidt, and L. E. Holmquist. Antonius: A mobile search engine for the physical world. In *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*, pages 179–182. ACM, 2013.

[63] M. Funk, S. Schneegass, M. Behringer, N. Henze, and A. Schmidt. An interactive curtain for media usage in the shower. In *Proceedings of the 4th International Symposium on Pervasive Displays*, pages 225–231. ACM, 2015.

[64] M. Funk, A. S. Shirazi, S. Mayer, L. Lischke, and A. Schmidt. Pick from here!: An interactive mobile cart using in-situ projection for order picking. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing*, pages 601–609. ACM, 2015.

[65] B. für Arbeit. Der arbeitsmarkt in deutschland–zeitarbeit–aktuelle entwicklungen, 2016.

[66] S. Gauglitz, C. Lee, M. Turk, and T. Höllerer. Integrating the physical environment into mobile remote collaboration. In *Proceedings of the 14th international conference on Human-computer interaction with mobile devices and services*, pages 241–250. ACM, 2012.

[67] N. Gavish, T. Gutiérrez, S. Webel, J. Rodríguez, M. Peveri, U. Bockholt, and F. Tecchia. Evaluating virtual reality and augmented reality training for industrial maintenance and assembly tasks. *Interactive Learning Environments*, (ahead-of-print):1–21, 2013.

[68] S. W. Greenwald, M. Funk, L. Loreti, D. Mayo, and P. Maes. Eva: Exploratory learning with virtual companions sharing attention and context. In *Advanced Learning Technologies (ICALT), 2016 IEEE 16th International Conference on*. IEEE, 2016.

[69] J. Grubert, D. Hamacher, R. Mecke, I. Böckelmann, L. Schega, A. Huckauf, M. Urbina, M. Schenk, F. Doil, and J. Tumler. Extended investigations of user-related issues in mobile industrial ar. In *Mixed and Augmented Reality (ISMAR), 2010 9th IEEE International Symposium on*, pages 229–230. IEEE, 2010.

[70] A. Guo, S. Raghu, X. Xie, S. Ismail, X. Luo, J. Simoneau, S. Gilliland, H. Baumann, C. Southern, and T. Starner. A comparison of order picking assisted by head-up display (hud), cart-mounted display (cmd), light, and paper pick list. In *Proceedings of the 2014 ACM International Symposium on Wearable Computers*, pages 71–78. ACM, 2014.

[71] A. Gupta, D. Fox, B. Curless, and M. Cohen. Duplotrack: a real-time system for authoring and guiding duplo block assembly. In *Proceedings of the 25th annual ACM symposium on User interface software and technology*, pages 389–402. ACM, 2012.

[72] P. Gurevich, J. Lanir, B. Cohen, and R. Stone. Teleadvisor: a versatile augmented reality tool for remote assistance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 619–622. ACM, 2012.

[73] J. Hahn, B. Ludwig, and C. Wolff. Augmented reality-based training of the pcb assembly process. In *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia*, pages 395–399. ACM, 2015.

[74] T. Haipeter and C. Slomka. Industriebeschäftigung im wandel: Arbeiter, angestellte und ihre arbeitsbedingungen. Technical report, SOEPpapers on Multidisciplinary Panel Data Research, 2014.

[75] J. Hardy and J. Alexander. Toolkit support for interactive projected displays. In *Proceedings of the 11th International Conference on Mobile and Ubiquitous Multimedia*, page 42. ACM, 2012.

[76] C. Harrison, H. Benko, and A. D. Wilson. Omnitouch: wearable multitouch interaction everywhere. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pages 441–450. ACM, 2011.

[77] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. *Advances in psychology*, 52:139–183, 1988.

[78] N. Hashimoto and I. Siio. Studi/o: positioning toy-block only by projection. In *Proceedings of the 2015 ACM International Joint Conference on Pervasive and Ubiquitous Computing and Proceedings of the 2015 ACM International Symposium on Wearable Computers*, pages 333–336. ACM, 2015.

[79] J. Heiser, D. Phan, M. Agrawala, B. Tversky, and P. Hanrahan. Identification and validation of cognitive design principles for automated generation of assembly instructions. In *Proceedings of the working conference on Advanced Visual Interfaces*, pages 311–319. ACM, 2004.

[80] S. Henderson and S. Feiner. Exploring the benefits of augmented reality documentation for maintenance and repair. *Visualization and Computer Graphics, IEEE Transactions on*, 17(10):1355–1368, 2011.

[81] S. J. Henderson and S. Feiner. Evaluating the benefits of augmented reality for task localization in maintenance of an armored personnel carrier turret. In *Mixed and Augmented Reality, 2009. ISMAR 2009. 8th IEEE International Symposium on*, pages 135–144. IEEE, 2009.

[82] S. J. Henderson and S. K. Feiner. Augmented reality in the psychomotor phase of a procedural task. In *Mixed and Augmented Reality (ISMAR), 2011 10th IEEE International Symposium on*, pages 191–200. IEEE, 2011.

[83] A. Hendricks. Un convention on the rights of persons with disabilities. *Eur. J. Health L.*, 14:273, 2007.

[84] H. Hua and C. Gao. Design of a bright polarized head-mounted projection display. *Applied optics*, 46(14):2600–2610, 2007.

[85] H. Ishii and B. Ullmer. Tangible bits: towards seamless interfaces between people, bits and atoms. In *Proceedings of the ACM SIGCHI Conference on Human factors in computing systems*, pages 234–241. ACM, 1997.

[86] R. J. Jacob, A. Girouard, L. M. Hirshfield, M. S. Horn, O. Shaer, E. T. Solovey, and J. Zigelbaum. Reality-based interaction: a framework for post-wimp interfaces. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 201–210. ACM, 2008.

[87] B. F. Johnson and A. J. Cuvo. Teaching mentally retarded adults to cook. *Behavior Modification*, 5(2):187–202, 1981.

[88] H. Kagermann, W.-D. Lukas, and W. Wahlster. Industrie 4.0: Mit dem internet der dinge auf dem weg zur 4. industriellen revolution. *VDI nachrichten*, 13:2011, 2011.

[89] J. Kim and H. Jun. Vision-based location positioning using augmented reality for indoor navigation. *Consumer Electronics, IEEE Transactions on*, 54(3):954–962, 2008.

[90] F. Klompmaker, K. Nebe, and A. Fast. dsensingni: a framework for advanced tangible interaction using a depth camera. In *Proceedings of the Sixth International Conference on Tangible, Embedded and Embodied Interaction*, pages 217–224. ACM, 2012.

[91] J. Knibbe, T. Grossman, and G. Fitzmaurice. Smart makerspace: An immersive instructional space for physical tasks. In *Proceedings of the 2015 International Conference on Interactive Tabletops & Surfaces*, pages 83–92. ACM, 2015.

[92] M. Kölz, D. Jordon, P. Kurtz, and T. Hörz. Adaptive assistance to support and promote performance-impaired people in manual assembly processes. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility*, pages 337–338. ACM, 2015.

[93] O. Korn, M. Funk, S. Abele, T. Hörz, and A. Schmidt. Context-aware assistive systems at the workplace: analyzing the effects of projection and gamification. In *Proceedings of the 7th International Conference on*

PErvasive Technologies Related to Assistive Environments, page 38. ACM, 2014.

[94] O. Korn, M. Funk, and A. Schmidt. Assistive systems for the workplace: Towards context-aware assistance. *Assistive Technologies for Physical and Cognitive Disabilities*, pages 121–133, 2015.

[95] O. Korn, M. Funk, and A. Schmidt. Design approaches for the gamification of production environments: a study focusing on acceptance. In *Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, page 6. ACM, 2015.

[96] O. Korn, M. Funk, and A. Schmidt. Towards a gamification of industrial production: a comparative study in sheltered work environments. In *Proceedings of the 7th ACM SIGCHI Symposium on Engineering Interactive Computing Systems*, pages 84–93. ACM, 2015.

[97] O. Korn, A. Schmidt, and T. Hörz. Assistive systems in production environments: exploring motion recognition and gamification. In *Proceedings of the 5th International Conference on PErvasive Technologies Related to Assistive Environments*. ACM, 2012.

[98] O. Korn, A. Schmidt, and T. Hörz. Augmented manufacturing: a study with impaired persons on assistive systems using in-situ projection. In *Proceedings of the 6th International Conference on PErvasive Technologies Related to Assistive Environments*, page 21. ACM, 2013.

[99] O. Korn, A. Schmidt, and T. Hörz. The potentials of in-situ-projection for augmented workplaces in production: a study with impaired persons. In *CHI'13 Extended Abstracts on Human Factors in Computing Systems*, pages 979–984. ACM, 2013.

[100] T. Kosch, R. Boldt, M. Hoppe, P. Knierim, and M. Funk. Exploring the optimal point of view in third person out-of-body experiences. In *Proceedings of the 9th ACM International Conference on PErvasive Technologies Related to Assistive Environments*, New York, NY, USA, 2016. ACM.

[101] T. Kosch, R. Kettner, M. Funk, and A. Schmidt. Comparing tactile, auditory, and visual assembly error-feedback for workers with cognitive impairments. In *Proceedings of the 18th international ACM SIGACCESS conference on Computers & accessibility*. ACM, 2016.

[102] M. Kritzler, M. Bäckman, A. Tenfält, and F. Michahelles. Wearable technology as a solution for workplace safety. In *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia*, pages 213–217. ACM, 2015.

[103] T. Kubitza and A. Schmidt. Towards a toolkit for the rapid creation of smart environments. In *End-User Development*, pages 230–235. Springer, 2015.

[104] G. E. Lancioni, M. F. O'Reilly, P. Seedhouse, F. Furniss, and B. Cunha. Promoting independent task performance by persons with severe developmental disabilities through a new computer-aided system. *Behavior Modification*, 24(5):700–718, 2000.

[105] X. Li, I. Y.-H. Chen, S. Thomas, and B. A. MacDonald. Using kinect for monitoring warehouse order picking operations. In *Proceedings of Australasian Conference on Robotics and Automation*, page 15, 2012.

[106] H. Lieberman. *Your wish is my command: Programming by example*. Morgan Kaufmann, 2001.

[107] N. Linder and P. Maes. Luminar: portable robotic augmented reality interface design and prototype. In *Adjunct proceedings of the 23nd annual ACM symposium on User interface software and technology*, pages 395–396. ACM, 2010.

[108] M. Löchtefeld, S. Gehring, R. Jung, and A. Krüger. guitar: supporting guitar learning through mobile projection. In *CHI'11 Extended Abstracts on Human Factors in Computing Systems*, pages 1447–1452. ACM, 2011.

[109] M. Löchtefeld, S. Gehring, J. Schöning, and A. Krüger. Shelftorchlight: Augmenting a shelf using a camera projector unit. In *Adjunct Proceedings of the Eighth International Conference on Pervasive Computing*, pages 1–4. Citeseer, 2010.

[110] I. S. MacKenzie and R. W. Soukoreff. Phrase sets for evaluating text entry techniques. In *CHI'03 extended abstracts on Human factors in computing systems*, pages 754–755. ACM, 2003.

[111] D. Marinos, B. Wöldecke, and C. Geiger. Prototyping natural interactions in virtual studio environments by demonstration: combining spatial mapping with gesture following. In *Proceedings of the Virtual Reality International Conference: Laval Virtual*, page 2. ACM, 2013.

[112] M. R. Marner, A. Irlitti, and B. H. Thomas. Improving procedural task performance with augmented reality annotations. In *Mixed and Augmented Reality (ISMAR), 2013 IEEE International Symposium on*, pages 39–48. IEEE, 2013.

[113] H. B. Maynard, G. J. Stegemerten, and J. L. Schwab. Methods-time measurement. 1948.

[114] G. McCalla, J. Greer, V. Kumar, P. Meagher, J. Collins, R. Tkatch, and B. Parkinson. A peer help system for workplace training. *B. d. Boulay, & R. Mizoguchi (Eds.), AI-ED*, 97(8):183–190, 1997.

[115] P. Milgram and F. Kishino. A taxonomy of mixed reality visual displays. *IEICE TRANSACTIONS on Information and Systems*, 77(12):1321–1329, 1994.

[116] B. A. Myers. Creating dynamic interaction techniques by demonstration. *ACM SIGCHI Bulletin*, 17(SI):271–278, 1986.

[117] N. Navab. Developing killer apps for industrial augmented reality. *Computer Graphics and Applications, IEEE*, 24(3):16–20, 2004.

[118] A. Nee, S. Ong, G. Chryssolouris, and D. Mourtzis. Augmented reality applications in design and manufacturing. *CIRP Annals-Manufacturing Technology*, 61(2):657–679, 2012.

[119] S. Nolle and G. Klinker. Augmented reality as a comparison tool in automotive industry. In *Proceedings of the 5th IEEE and ACM International Symposium on Mixed and Augmented Reality*, pages 249–250. IEEE Computer Society, 2006.

[120] W. H. Organization et al. World report on disability 2011. 2011.

[121] S. Ota, Y. Takegawa, T. Terada, and M. Tsukamoto. A method for wearable projector selection that considers the viewability of projected images. *Computers in Entertainment (CIE)*, 8(3):17, 2010.

[122] M. Otto, M. Prieur, and E. Rukzio. Using scalable, interactive floor projection for production planning scenario. In *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces*, pages 363–368. ACM, 2014.

[123] V. Paelke. Augmented reality in the smart factory: Supporting workers in an industry 4.0. environment. In *Emerging Technology and Factory Automation (ETFA), 2014 IEEE*, pages 1–4. IEEE, 2014.

[124] B. Petzold, M. F. Zaeh, B. Faerber, B. Deml, H. Egermeier, J. Schilp, and S. Clarke. A study on visual, auditory, and haptic feedback for assembly tasks. *Presence*, 13(1):16–21, 2004.

[125] K. L. Pierce and L. Schreibman. Teaching daily living skills to children with autism in unsupervised settings through pictorial self-management. *Journal of applied behavior analysis*, 27(3):471–481, 1994.

[126] C. Pinhanez. The everywhere displays projector: A device to create ubiquitous graphical interfaces. In *Ubicomp 2001: Ubiquitous Computing*, pages 315–331. Springer, 2001.

[127] O. Pötzsch. Geburten in deutschland. *Statistisches Bundesamt: Wiesbaden*, 2012.

[128] R. Rädle, H.-C. Jetter, N. Marquardt, H. Reiterer, and Y. Rogers. Huddlelamp: Spatially-aware mobile displays for ad-hoc around-the-table collaboration. In *Proceedings of the Ninth ACM International Conference on Interactive Tabletops and Surfaces*, pages 45–54. ACM, 2014.

[129] R. Raskar, J. Van Baar, P. Beardsley, T. Willwacher, S. Rao, and C. Forlines. ilamps: geometrically aware and self-configuring projectors. In *ACM SIGGRAPH 2006 Courses*, page 7. ACM, 2006.

[130] R. Raskar, G. Welch, and H. Fuchs. Spatially augmented reality. In *First IEEE Workshop on Augmented Reality (IWAR'98)*, pages 11–20. Citeseer, 1998.

[131] M. Rauterberg and E. Styger. Positive effects of sound feedback during the operation of a plant simulator. In *Human-Computer Interaction*, pages 35–44. Springer, 1994.

[132] H. Regenbrecht, G. Baratoff, and W. Wilke. Augmented reality projects in the automotive and aerospace industries. *Computer Graphics and Applications, IEEE*, 25(6):48–56, 2005.

[133] R. Reif, W. A. Günthner, B. Schwerdtfeger, and G. Klinker. Pick-by-vision comes on age: evaluation of an augmented reality supported picking system in a real storage environment. In *Proceedings of the 6th International Conference on Computer Graphics, Virtual Reality, Visualisation and Interaction in Africa*, pages 23–31. ACM, 2009.

[134] P. Richard, G. Burdea, D. Gomez, and P. Coiffet. A comparison of haptic, visual and auditive force feedback for deformable virtual objects. In

Proceedings of the Internation Conference on Automation Technology (ICAT), pages 49–62, 1994.

[135] E. Rukzio, P. Holleis, and H. Gellersen. Personal projectors for pervasive computing. *IEEE Pervasive Computing*, 11(2):30–37, 2012.

[136] S. Rüther, T. Hermann, M. Mracek, S. Kopp, and J. Steil. An assistance system for guiding workers in central sterilization supply departments. In *Proceedings of the 6th International Conference on PErvasive Technologies Related to Assistive Environments*, page 3. ACM, 2013.

[137] N. Sakata, T. Kurata, T. Kato, M. Kourogi, and H. Kuzuoka. Wacl: Supporting telecommunications using wearable active camera with laser pointer. In *Proceedings of the 7th IEEE International Symposium on Wearable Computers*, ISWC '03, pages 53–, Washington, DC, USA, 2003. IEEE Computer Society.

[138] N. Sakata, T. Kurata, and H. Kuzuoka. Visual assist with a laser pointer and wearable display for remote collaboration. In *Proceedings of 2nd international conference on Collaboration Technologies*, 2006.

[139] A. L. Sauer, A. Parks, and P. C. Heyn. Assistive technology effects on the employment outcomes for people with cognitive disabilities: a systematic review. *Disability & Rehabilitation: Assistive Technology*, 5(6):377–391, 2010.

[140] T. Schlömer, B. Poppinga, N. Henze, and S. Boll. Gesture recognition with a wii controller. In *Proceedings of the 2nd international conference on Tangible and embedded interaction*, pages 11–14. ACM, 2008.

[141] A. Schmidt, M. Beigl, and H.-W. Gellersen. There is more to context than location. *Computers & Graphics*, 23(6):893–901, 1999.

[142] B. Schwerdtfeger, D. Pustka, A. Hofhauser, and G. Klinker. Using laser projectors for augmented reality. In *Proceedings of the 2008 ACM symposium on Virtual reality software and technology*, pages 134–137. ACM, 2008.

[143] B. Schwerdtfeger, R. Reif, W. A. Gunthner, G. Klinker, D. Hamacher, L. Schega, I. Bockelmann, F. Doil, and J. Tumler. Pick-by-vision: A first stress test. In *Mixed and Augmented Reality, 2009. ISMAR 2009. 8th IEEE International Symposium on*, pages 115–124. IEEE, 2009.

[144] A. S. Shirazi, M. Funk, F. Pfleiderer, H. Glück, and A. Schmidt. Mediabrain: Annotating videos based on brain-computer interaction. In *Mensch & Computer*, pages 263–272, 2012.

[145] B. Shneiderman. *Designing the user interface: strategies for effective human-computer interaction*, volume 3.

[146] S. E. Steed and J. R. Lutzker. Using picture prompts to teach an adult with developmental disabilities to independently complete vocational tasks. *Journal of Developmental and Physical Disabilities*, 9(2):117–133, 1997.

[147] I. E. Sutherland. A head-mounted three dimensional display. In *Proceedings of the December 9-11, 1968, fall joint computer conference, part I*, pages 757–764. ACM, 1968.

[148] G. Suzuki, T. Murase, and Y. Fujii. Projecting recorded expert hands at real size, at real speed, and onto real objects for manual work. In *Companion Publication of the 21st International Conference on Intelligent User Interfaces*, pages 13–17. ACM, 2016.

[149] A. Tang, C. Owen, F. Biocca, and W. Mou. Comparative effectiveness of augmented reality in object assembly. In *Proceedings of the SIGCHI conference on Human factors in computing systems*, pages 73–80. ACM, 2003.

[150] J. Tumler, F. Doil, R. Mecke, G. Paul, M. Schenk, E. A. Pfister, A. Huckauf, I. Bockelmann, and A. Roggentin. Mobile augmented reality in industrial applications: Approaches for solution of user-related issues. In *Proceedings of the 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pages 87–90. IEEE Computer Society, 2008.

[151] J. A. Ward, P. Lukowicz, G. Troster, and T. E. Starner. Activity recognition of assembly tasks using body-worn microphones and accelerometers. *Pattern Analysis and Machine Intelligence, IEEE Transactions on*, 28(10):1553–1567, 2006.

[152] K. A. Weaver, H. Baumann, T. Starner, H. Iben, and M. Lawo. An empirical task analysis of warehouse order picking using head-mounted displays. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, pages 1695–1704. ACM, 2010.

[153] M. Weing, A. Röhlig, K. Rogers, J. Gugenheimer, F. Schaub, B. Könings, E. Rukzio, and M. Weber. Piano: enhancing instrument learning via

interactive projected augmentation. In *Proceedings of the 2013 ACM conference on Pervasive and ubiquitous computing adjunct publication*, pages 75–78. ACM, 2013.

[154] M. Weiser. The computer for the 21st century. *Scientific american*, 265(3):94–104, 1991.

[155] H.-P. Wiendahl, J. Reichardt, and P. Nyhuis. *Handbook Factory Planning and Design*. (pp. 150–153) Springer, 2015.

[156] K. D. Willis, I. Poupyrev, S. E. Hudson, and M. Mahler. Sidebyside: ad-hoc multi-user interaction with handheld projectors. In *Proceedings of the 24th annual ACM symposium on User interface software and technology*, pages 431–440. ACM, 2011.

[157] A. D. Wilson. Touchlight: an imaging touch screen and display for gesture-based interaction. In *Proceedings of the 6th international conference on Multimodal interfaces*, pages 69–76. ACM, 2004.

[158] A. D. Wilson. Using a depth camera as a touch sensor. In *ACM international conference on interactive tabletops and surfaces*, pages 69–72. ACM, 2010.

[159] C. Winkler, J. Seifert, D. Dobbelstein, and E. Rukzio. Pervasive information through constant personal projection: the ambient mobile pervasive display (amp-d). In *Proceedings of the 32nd annual ACM conference on Human factors in computing systems*, pages 4117–4126. ACM, 2014.

[160] W. Wohlgemuth and G. Triebfürst. Arvika: augmented reality for development, production and service. In *Proceedings of DARE 2000 on Designing augmented reality environments*, pages 151–152. ACM, 2000.

[161] K. Wolf, M. Funk, P. Knierim, and M. Löchtefeld. Survey of interactive displays through mobile projections. *International Journal of Mobile Human Computer Interaction*, 8(4):30–43, 2016.

[162] K. Wolf, S. Schneegass, N. Henze, D. Weber, V. Schwind, P. Knierim, S. Mayer, T. Dingler, Y. Abdelrahman, T. Kubitza, et al. Tuis in the large: Using paper tangibles with mobile devices. In *Proceedings of the 33rd Annual ACM Conference Extended Abstracts on Human Factors in Computing Systems*, pages 1579–1584. ACM, 2015.

[163] M. Yuan, S. Ong, and A. Nee. Augmented reality for assembly guidance using a virtual interactive tool. *International Journal of Production Research*, 46(7):1745–1767, 2008.

[164] J. Zauner, M. Haller, A. Brandl, and W. Hartman. Authoring of a mixed reality assembly instructor for hierarchical structures. In *Mixed and Augmented Reality, 2003. Proceedings. The Second IEEE and ACM International Symposium on*, pages 237–246. IEEE, 2003.

[165] X. S. Zheng, C. Foucault, P. Matos da Silva, S. Dasari, T. Yang, and S. Goose. Eye-wearable technology for machine maintenance: Effects of display position and hands-free operation. In *Proceedings of the 33rd Annual ACM Conference on Human Factors in Computing Systems*, pages 2125–2134. ACM, 2015.

[166] J. Zhou, I. Lee, B. Thomas, R. Menassa, A. Farrant, and A. Sansome. Applying spatial augmented reality to facilitate in-situ support for automotive spot welding inspection. In *Proceedings of the 10th International Conference on Virtual Reality Continuum and Its Applications in Industry*, pages 195–200. ACM, 2011.

VII

Appendix

Markus Funk

Augustenstraße 81
70178 Stuttgart
Germany

Phone:	+49 711 685-60092
Email:	makufunk@hotmail.com
Homepage:	www.makufunk.de
Birthday:	07/25/1987
Nationality:	German

Curriculum Vitae

Markus Funk is currently finishing his PhD at the Human-Computer Interaction group of Albrecht Schmidt at the Institute for Visualization and Interactive Systems at the University of Stuttgart. His research interests encompass Augmented Reality, Virtual Reality, and Human-Drone Interaction.

Education and training

University of Stuttgart 02/2013 - present
Research Associate and *PhD-student*
Human-Computer Interaction group
Supervisor: Albrecht Schmidt

MIT Media Lab 04/2016 – 06/2016
Visiting Researcher
Fluid Interfaces
Supervisor: Pattie Maes

Yahoo! Labs, Sunnyvale, CA, USA 06/2012 – 09/2012
Research Intern
Supervisor: Lars Erik Holmquist

Diploma: Software Engineering 10/2007 – 12/2012
Major: Architecture of Application Systems
University of Stuttgart

University of Stuttgart 03/2011 – 06/2012
Student Research Assistant
Human-Computer Interaction group

Intergraph PP&M Germany in Süßen 10/2009 - 04/2010
Software Developer (C++)

Technical skills

Project management experience in software projects:
- motionEAP: 4 developers (runtime 4 years)
- Mobi Track: 12 developers (runtime 1 year)

Programming: Java, C#, C++, MySQL, Android

Teaching experience:

- **Introduction to Human-Computer Interaction**
  Summer term 2016
  (with Prof. Dr. Albrecht Schmidt & Dr. Tonja Machulla)
- **Empirical Methods for Media Informatics**
  Winter term 2015 (with Dr. Lewis Chuang)
- **Practical Project - WearableOS**
  Summer term 2014 + winter term 2014
  (with Dr. Stefan Schneegass)
- **Practical Project – Keepsake**
  Summer term 2013 + winter term 2013
  (with Prof. Dr. Niels Henze and Prof. Dr. Florian Alt)
- **Introduction to Human-Computer Interaction**
  Summer term 2013
  (with Prof. Dr. Albrecht Schmidt and Prof. Dr. Niels Henze)

Further Markus Funk was supervising 12 bachelor theses and 10 master theses.

Reviewing Activities

Conference on Human Factors in Computing Systems (CHI)	2014-16
Augmented Human International Conference (AH)	2014
Nordic Conference on Human-Computer Interaction (NordiCHI)	2014, 2016
International Conference on the Internet of Things (IOT)	2015-16
GermanHCI - Mensch und Computer	2015-16
International Conference on Mobile and Ubiquitous Multimedia (MUM)	2015
International Symposium on Pervasive Displays (PerDis)	2015
International conference on Tangible, Embedded and Embodied Interaction (TEI)	2016
ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp)	2016
International Conference on Human-Computer Interaction with Mobile Devices and Services (mobileHCI)	2016
International Symposium on Wearable Computers (ISWC)	2016

Markus Funk

Augmented Reality at the Workplace:
A Context-Aware Assistive System using In-Situ Projection

Augmented Reality has been used for providing assistance during manual assembly tasks for more than 20 years. Due to recent improvements in sensor technology, creating context-aware Augmented Reality systems, which can detect interaction accurately, becomes possible. Additionally, the increasing amount of variants of assembled products and being able to manufacture ordered products on demand, leads to an increasing complexity for assembly tasks at industrial assembly workplaces. The resulting need for cognitive support at workplaces and the availability of robust technology enables us to address real problems by using context-aware Augmented Reality to support workers during assembly tasks.

In this thesis, we explore how assistive technology can be used for cognitively supporting workers in manufacturing scenarios. By following a user-centered design process, we identify key requirements for assistive systems for both continuously supporting workers and teaching assembly steps to workers. Thereby, we analyzed three different user groups: inexperienced workers, experienced workers, and workers with cognitive impairments. Based on the identified requirements, we design a general concept for providing cognitive assistance at workplaces which we apply to the area of manual assembly and order picking.

This thesis broadens the understanding of how in-situ projection can be used at manual assembly workplaces and order picking scenarios. It outlines challenges and opportunities of in-situ projection technology and compares it to other state-of-the-art instruction systems. The results of this thesis reveal that in-situ projection is a promising technology that is soon robust enough to be used for training new workers and continuously supporting workers with cognitive impairments.