Augmented Reality Training for Industrial Assembly Work – Are Projection-based AR Assistive Systems an Appropriate Tool for Assembly Training?

Sebastian Büttner
OWL University of Applied Sciences, Lemgo & Clausthal University of Technology, Clausthal-Zellerfeld, Germany
sebastian.buettner@th-owl.de

Michael Prilla
Clausthal University of Technology, Clausthal-Zellerfeld, Germany
michael.prilla@tu-clausthal.de

Carsten Röcker
OWL University of Applied Sciences, Lemgo & Fraunhofer IOSB-INA, Lemgo, Germany
carsten.roecker@th-owl.de

ABSTRACT

Augmented Reality (AR) systems are on their way to industrial application, e.g. projection-based AR is used to enhance assembly work. Previous studies showed advantages of the systems in permanent-use scenarios, such as faster assembly times. In this paper, we investigate whether such systems are suitable for training purposes. Within an experiment, we observed the training with a projection-based AR system over multiple sessions and compared it with a personal training and a paper manual training. Our study shows that projection-based AR systems offer only small benefits in the training scenario. While a systematic mislearning of content is prevented through immediate feedback, our results show that the AR training does not reach the personal training in terms of speed and recall precision after 24 hours. Furthermore, we show that once an assembly task is properly trained, there are no differences in the long-term recall precision, regardless of the training method.

Author Keywords

Industrial Augmented Reality; Projection-based Augmented Reality; Assembly; Training; Assistive System; Empirical Study; Experiment

CCS Concepts

•**Human-centered computing → Mixed / augmented reality; Empirical studies in HCI;**

INTRODUCTION

Despite an ever increasing degree of automation, it is predicted that manual industrial work will not disappear in future [14]. Due to fluctuating demand, companies are regularly forced to train (partly unskilled) workers in a short time. This observation raises questions to the HCI community on how to support industrial training with new interaction technologies.

CHI'20, April 25–30, 2020, Honolulu, HI, USA

ISBN 978-1-4503-6708-0/20/04. . . 15.00

DOI: https://doi.org/10.1145/3313831.3376720

In the context of industrial assembly, a lot of assistive systems have been presented which provide instructions at the workplace. Most of the systems create Augmented Reality (AR) environments, either using head-mounted displays (HMDs) [30] or in-situ projections [19], that integrate virtual information about the next assembly step or about errors into the physical environment. Short-term studies show that AR assistive systems can reduce task completion time and errors of procedural tasks [4, 27]. Even though the results of these studies are promising, it is questionable whether such systems are to be used permanently and in the long run (*permanent-use scenario*), which only makes sense if an assembly work place does not contain any repetitive activities. Even worse, long-term studies of such systems show negative effects if applied beyond the training phase [17]. Consequently, the permanent-use scenario is questionable.

In this paper, we focus on training, which has not been considered in depth in earlier research: could the proposed projection-based AR systems be used temporarily for an efficient on-the-job training (*training scenario*)? This paper contributes to the research on assitive systems by presenting an experiment that investigates training with projection-based AR systems in terms of two possible effects: First, *training efficiency* in terms of training time; second, how well a learned task is remembered, which we will call *knowledge sustainability*. We compare AR training with personal training (*conventional case*, trainer required) and training with a paper manual (*low-cost case*, no trainer required).

RELATED WORK

Using AR to enhance industrial work has been a research topic for years. Fite-Georgel coined the term *Industrial AR (IAR)* [15] that we will use in the following to describe AR systems that are specifically designed for the support of industrial work. This section sums up the current state of the development and evaluation of AR assistive systems in general, followed by a view on IAR-based training.

AR Assistive Systems

A lot of concepts have been created that should exploit the full potential of IAR. One application area is the support of manual assembly through assistive systems. According to Fellmann et al. [14] we understand an assistive system as “a context-aware system [...] that supports a user with the execution of a task and adapts depending on the progress of the task.” Previous assistive systems for assembly make use of mobile devices, HMDs or in-situ projections, using video projectors or lasers (compare [7, 15, 31]).

An early work investigating the effectiveness of AR assistive systems for assembly tasks was done by Tang et al. [30], who compared the use of paper instructions, computer instructions on an LCD screen, computer instructions on an HMD and spatially registered instructions on an HMD. In their experiment, participants had to assemble a 56-step task consisting of Lego Duplo blocks. They measured the task completion time and number of errors and found out that spatial AR instructions do not have advantages regarding task completion time, but that they can reduce the error rate [30]. In line with the results are the findings of Zheng et al. who investigated procedural maintenance scenarios with HMDs and confirmed that the use of HMDs does not lead to faster task completion times [35]. Billinghurst et al. [3] investigated assitive systems based on mobile devices. In their user study, they did not compare their system to a (traditional) baseline. Rather they evaluate different AR and VR views with a 3-D puzzle task. By measuring the task completion time, they showed that participants performed best under animated AR instructions [3]. A similar system has been presented by Gorecky et al., who used tablet computers [22]. A comprehensive overview on AR usability studies is given by Dey et al. [11], who provide a systematic review on research done from 2005–2014 including 30 papers of user studies of industrial applications. Most of the studies listed focus on task completion time and error rate of various supported work processes. Their review contains mainly AR based on HMDs or mobile devices. Only a very small number of papers (and none about industrial applications) used in-situ projections.

Considering in-situ projections, Schwerdtfeger et al. proposed the use of laser projectors to provide instructions in the work space of the user [29]. By developing a stationary and a head-mounted prototype, they show the technological feasibility of the systems. A comprehensive investigation with regard to the impacts on work performance in a sheltered work scenario has been done by Funk et al. with a stationary prototype of an AR assistive systems with in-situ projections created by video projectors. In [16] they analyze different visualizations for such systems and show that spatial contour visualizations help to reduce mental load. In [21] they measure task completion time and error rate of a Lego Duplo task and show that assembly times are faster and fewer errors are produced, when in-situ projections are used. Furthermore, impaired workers are able to perform more complex tasks with the help of the assistive system. In the context of in-situ projection gamification approaches [25] and the integration of physiological data [12] have been described. In addition to stationary systems, systems with mobile projectors were also presented [8, 20].

Further studies aimed at comparing different types of AR assistive systems: Büttner et al. compare assembly support with HMD-based AR and projection-based AR with a paper manual as a baseline measuring task completion time and error rate, indicating a weak performance of the HMD-based AR system [6]. Their study also analyzes qualitative data that explains the particular issues with HMD-based AR. In the same year, Funk et al. [19] compare assembly support with an (centrally positioned) HMD-based view, projection-based AR, AR using a tablet and a paper baseline. The results of their study are in line with the results of Büttner et al. [6], but are more detailed regarding the measured times. They show that the sub-tasks of locating parts and assembling are significantly faster with in-situ projections [19]. A study of Marner et al. [27] comparing in-situ projections with instructions on a display shows that projection-based AR leads to faster task completion times and fewer errors when executing procedural tasks. While previous work considers a short-term use of the assistive systems, Funk et al. presented a first long-term study for projection-based AR assistive systems for different user groups (untrained vs. expert) in the field [17]. They emphasize that “in-situ instructions are useful during the learning phase” for untrained workers [17]. However, they also observe that the task completion time and the perceived cognitive load increases, if the system is used by the experts or if it is used beyond the training phase of the untrained users.

Based on the previous works described above, we identified training as the most interesting and practically relevant use case for projection-based AR assitive systems. While most of the studies above considered task completion time, errors and perceived performance (usually measured by NASA-TLX [23]) to evaluate certain systems, we argue that, if projection-based AR assistive systems should be used for training, we need to focus on different effects, namely training efficiency and knowledge sustainability.

Training with AR

There are few publications that focus on the training with AR assistive systems: Boud et al. investigated AR and virtual reality (VR) as a training tool for industrial assembly tasks [4]. However, in their study they also measured task completion time of the participants and did not look into training efficiency or sustainability. Similarly, Baird et al. investigated AR as a training tool for motherboard assembly, also focusing on task completion time and error rate [2]. More recently, Westerfield et al. investigated AR explicitly for learning assembly tasks [32, 33]. They conducted a study in which users had to assemble computer components on a motherboard and measured the knowledge of the participants with pre- and post-tests. However, they compared two different versions of an AR assitive system: one version with standardized instructions and one adaptive version with individual customized instructions for each participant. While showing the effectiveness of the individual training, they did not collect data of a classical training to evaluate the AR training in general. A study similar to the study presented here (but with different control conditions) was conducted by Wilschut et al. [34], who compared how users learn new assembly tasks with in-situ projections compared to electronic working instructions (EWI) on a stationary display. Furthermore, they investigated how chunking of instructions influences the learning process. In their study, they did not find any differences between AR and

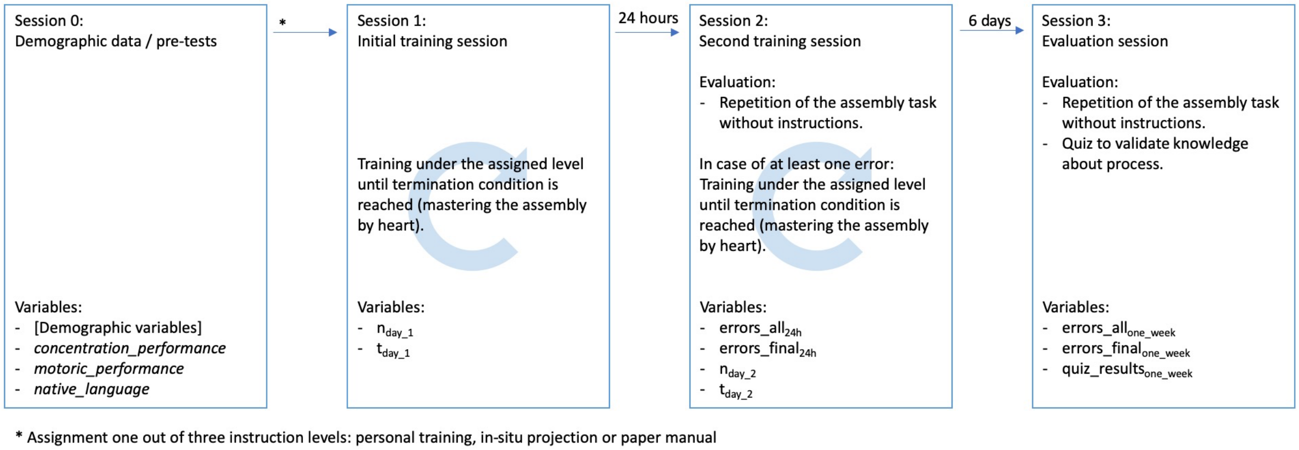

Figure 1. Overview over the procedure and the data collected per session.

EWI instructions in terms of training efficiency, error rates or perceived workload. Some further industrial AR training systems have been proposed in the context of vocational training of machine maintenance [1, 13, 28] but without a deeper empirical analysis of the effects on training performance.

To sum up our analysis on related work, we can say that, even if there have been a lot of systems presented that should facilitate the training by means of AR systems, there is a weak empirical evidence about the impacts of these systems on the training efficiency and sustainability.

HYPOTHESES

Based on the presented literature, we define our hypothesis. We assume that assembly tasks are learned in the fastest way with a personal training. Based on the results of Funk et al. [17] (who investigated task completion time), we assume that people can learn faster with an AR assitive system compared to a paper manual. However, we assume, that we cannot reach the training efficiency of a personal training, since a trainer can individually respond to trainees. Regarding knowledge sustainability, we assume that spatially presented information lead to better memorability compared to images in a paper manual. However, we do not assume that the personal training is superior. Consequently, we want to investigate the following hypotheses (H1 to H3 relate to training efficiency, H4 to H6 relate to knowledge sustainability):

H_1: Assembly tasks are learned faster with personal training than with a paper manual.

H_2: Assembly tasks are learned faster with a projection-based AR assistive system than with a paper manual.

H_3: Assembly tasks are learned faster with personal training than with a projection-based AR assistive system.

H_4: An assembly task conveyed by personal training is better remembered than one conveyed by a paper manual.

H_5: An assembly task conveyed through a projection-based AR assistive system is better remembered than one conveyed by a paper manual.

H_6: An assembly task is remembered equally well, regardless if it is conveyed through a projection-based AR assistive system or by personal training.

METHODOLOGY

In order to find out about how people learn with projection-based AR systems, we carried out an experiment and observed and measured how people learn under different instruction conditions. This section describes how our experiment was set up and carried out.

Design

The experiment followed a between-subjects design with the independent variable *instruction_method* having three levels: personal training, in-situ projection and paper manual. Each participant had to take part in four sessions: one pre-test session, two training sessions and one evaluation session. An overview of the procedure is shown in Figure 1.

As dependent variables, we collected the following data:

n_{day_x}	Number of training cycles during training session x, $x \in \{1,2\}$.
t_{day_x}	Training time during training session x, $x \in \{1,2\}$.
$errors\_all_{24h}$	Number of all errors occurring during the first assembly cycle in training session 2 (24 hours after the first training session).
$errors\_final_{24h}$	Number of errors that had an impact on the final result of the first assembly cycle in training session 2.
$errors\_all_{7d}$	Number of all errors occurring during the first assembly cycle in evaluation session 3 (one week after training session 1).
$errors\_final_{7d}$	Number of errors that had an impact on the final result of the first assembly cycle in evaluation session 3.
$quiz\_results_{7d}$	Number of correct questions in multiple-choice test in evaluation session 3.

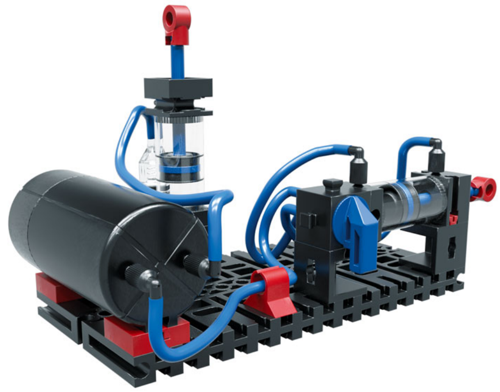

Figure 2. Final result of the assembly task.
(Image source: fischertechnik, ©fischertechnik GmbH)

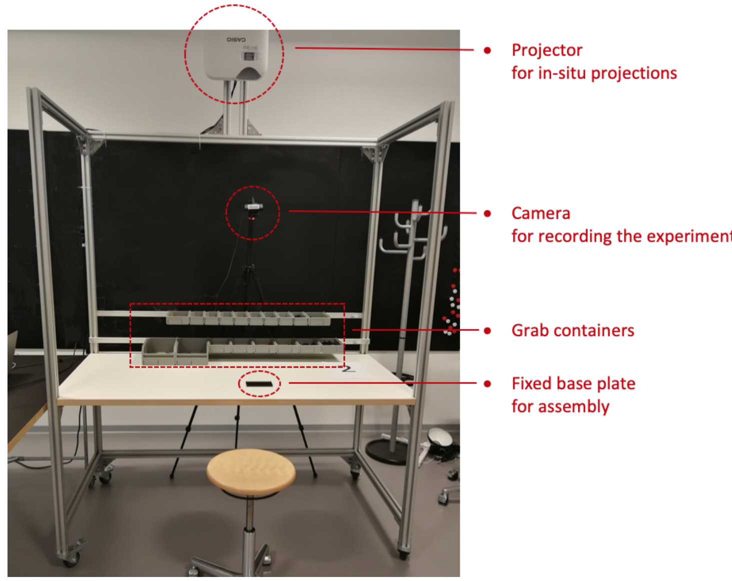

Figure 3. Assembly workplace for the study.

Participants

For our experiment, we recruited 33 participants in two separate sets. The first part of the study contained 22 students from two courses at our university who participated on a voluntary basis. The participants of the first set did not receive any payment for their participation, but we raffled a Tronico metal construction kit among the participants. For the second part, we recruited 11 more people by offering a financial incentive (25 euros) for taking part in the study. The background of these 11 people is more diverse than that of the first set: the majority of them is from the university community (employees, interns, students from other faculties, etc.), but there was also one participant that was not related to the university.

Out of the 33 participants, 27 completed the study. Six participants – all from the first set – dropped out because they did not show up to all sessions. The data of three more participants was excluded from the analysis because the three participants did not learn the task correctly (see section 'Results and Discussion'). From the remaining 24 participants, 7 were assigned to the paper condition (set 1: 4, set 2: 3), 9 to the AR condition (set 1: 5, set 2: 4) and 8 to the personal training condition (set 1: 5, set 2: 3). In the following, we present the demographics of the remaining participants.

Of the 24 participants, 20 were male and 4 were female. The age ranges from 14 to 51 years (M = 25.17 yr., SD = 7,79 yr.). 18 of the participants had German as native language, whereas 6 had other native languages. (The experiment was carried out in German.) 22 are right-handed and 2 are left-handed.

The pre-tests showed that our participants had a concentration performance value (KL) ranging from 65 to 241 (M = 159.50, SD = 36.19), which corresponds to the age group '20–39 years' (M = 158.6, SD = 29.4) in the calibration test of Brickenkamp et al. [5]. Consequently, we assume an average concentration performance of our sample. Regarding the motor performance of our participants, we measured an average task completion time of the general assembly task (GAT) of 164.38 seconds (SD = 36.19 sec.), which is almost the value reported by Funk et al. (M = 172.48 sec.) [18], so we assume average motor skills in our sample.

Apparatus

Here we describe the task, the assembly workplace, the particular materials for the three instruction levels (the developed AR assistive system, the paper manual and the trainer guidelines) as well as the multiple-choice quiz for the evaluation.

Task

As a basis for our assembly task, we chose a commercial construction kit from the manufacturer 'fischertechnik' (set: 540946 – 'Pneumatics Beginner', model: 'pneumatic power functional model'), which is shown in Figure 2.[1] This particular model consists of 26 single parts, including a base plate, two pneumatic cylinders, two hand valves, one air reservoir and four flexible tubes. For our task, we preassembled the air reservoir with two other parts (front holder), resulting in a total of 24 parts to be assembled in 23 steps. The base plate was fixed to the assembly table with double-sided adhesive tape for all three conditions. Furthermore, we defined a sequence in which parts have to be assembled. While some parts have a natural assembly sequence, the defined sequence has been arbitrarily chosen in some places, mainly based on the principle of locality.

Compared to previous studies that often used LEGO Duplo bricks (e.g. [18, 24, 26, 30]), the fischertechnik task used in this study adds some complexity, since different parts have to be assembled in a different manner. However, the task can still be performed without tools, which makes an introduction to the tools unnecessary, eliminates potential influence of previous tool usage and removes the risk of injury during the study.

[1] see https://www.fischertechnik.de/en/products/teaching/stem-kits/540946-pneumatics-beginner

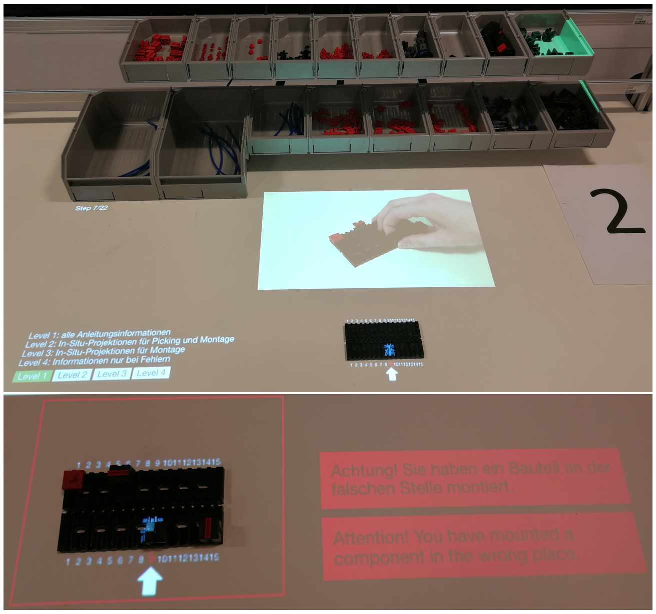

Figure 4. Projection-based user interface showing (a) the next assembly step (level 1) and (b) indicating an assembly error.

Assembly Workplace

The assembly workplace setup is shown in Figure 3. This setup was used for all conditions. The setup is comparable to industrial manual assembly workstations and similar to previous studies (see 'Related work'): the participants sat on a chair in front of the work surface and were able to pick all required parts from grab containers in front of them. A camera mounted on a tripod covering the work space and the grab containers was used to record all sessions. On the left side of the workplace, the study director and an assistant sat rotated by 90 degrees, who were able to observe the work surface live and via a computer screen (top-view image of the camera above the work surface). For the personal training condition, an additional trainer sat next to the participants in front of the workplace.

AR Assistive System

We set up a projection-based AR assistive system that mimics previous research or commercial systems (including our own system presented in [9]). For the in-situ projections, we used a Casio projector (model: 'XJ-V110W') with 3,500 ANSI lumen mounted on top of the structure of the assembly station. This projector covered the complete work surface including the grab containers.

The user interface of the assitive system was developed in HTML5. It contained the following elements (see Figure 4a): picking information that highlighted the box that contained the parts to be picked next in a green color; a video showing the current assembly step above the base plate in an infinite loop; animated in-situ projections on the base plate to show mounting points (if parts had to be mounted to the base plate an additional numbering scheme was shown); and some buttons to the bottom-left that could be used by the participant to select different support levels. Error information indicated by highlighting either the grab containers in red color if wrong parts had been picked or by showing a red rectangle around the base plate including an error message describing what exactly went wrong (see Figure 4b).

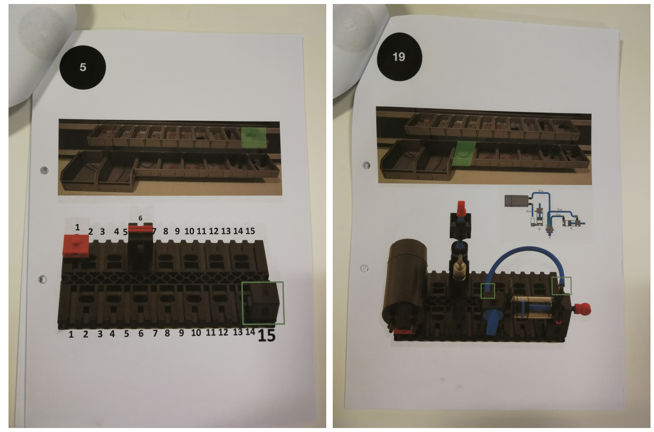

Figure 5. Examples of pages of the paper manual.

There were four different support levels implemented into the system, which could be freely chosen by the participant and changed at any time. Level 1 contained all information described above. Level 2 contained the information of level 1 except the video. Level 3 contained the information of level 2 except the picking information. And level 4 ('silent mode') did not give any assembly information except that it notified the user on errors.

Even though we have introduced assistive systems with hand-tracking in our previous work [9], we decided to recognize the state of the user with a Wizard-of-Oz approach in this experiment. This decision was based on our experience regarding the technical reliability of the hand-tracking module during study situations. In particular, we had previously had issues with participants who picked multiple parts at the same time (which cannot be recognized by the hand-tracking module). This later led to unwanted displayed errors (due to the missing pick gesture). So instead of using the hand-tracking module, we implemented a control software in HTML5 that was used by the study assistant to switch to the next assembly step, to indicate errors to the participants if necessary and to change the level if the participant pressed one of the level buttons.

Paper Manual

The paper manual consisted of 26 pages. The first page contained the base plate with the numbering scheme. Each of the following pages contained one of the 23 steps; the last step (assembly of a tube with tube holder on base plate) was presented in three sub-steps on three pages due to the complexity of the step. The pages representing a step contained the picking information on the top by showing an image of the grab containers, in which the relevant container was highlighted in green (see Figure 5). Below, the current state including the part to be mounted was shown. The current part was furthermore highlighted with a green rectangle. If parts had to be assembled onto the base plate, the numbering scheme of the base plate was included (see Figure 5 left). If tubes had to be connected, an additional plan for the connections was included (see Figure 5 right).

Trainer Guidelines

The trainer in our experiment was the same person for all sessions. Even though the training itself was very free and the content of the explanations was influenced by the questions of the participants, we defined some guidelines that the trainer had to follow. The training always started with a first round, in which the trainer led the way step by step and the participant imitated each step directly. In the following rounds, the degree of assistance depended on the wishes of the respective study participant. When participants felt confident enough, the trainer observed and helped only with single steps; later the trainer observed and always pointed out errors. After each round, the trainer asked whether he should give the same amount or more or less help. Of course, the trainer had to reply to all questions relating to the task which came up during the training sessions.

Multiple-Choice Quiz

In order to evaluate the knowledge of the participants after one week, we developed an image-based multiple-choice quiz in HTML5. The quiz contained 25 questions of two types (see Figure 6). The first 17 questions mainly dealt with positions or orientations of parts. For each question of this type, two images were shown: one correct and one wrong stage of the assembly process. The participants had to select the one that seemed to be correct to them. The last 8 questions concerned the order of parts. A sequence of parts to be mounted was shown and participants had to click on one out of three parts which they thought was the next one in the sequence. The mouse clicks of the participants were automatically logged by the web application.

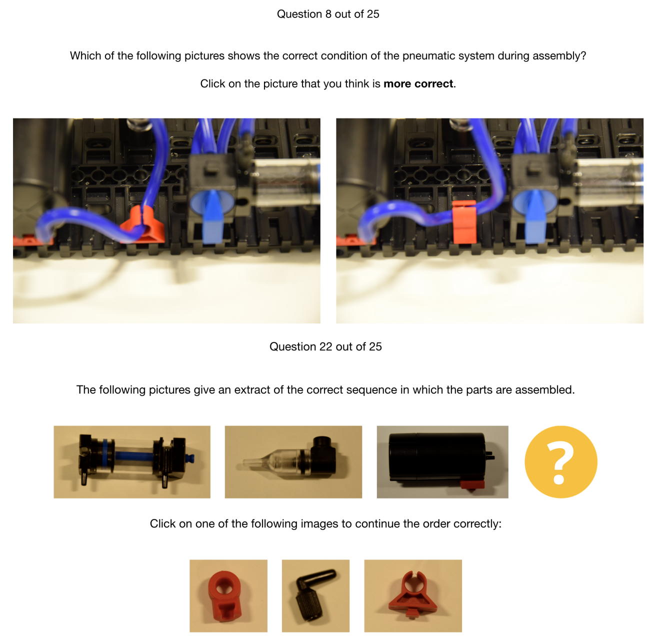

Figure 6. Screenshots of the quiz app showing the different question types in the multiple-choice quiz: (top) position and orientation questions (bottom) sequence questions. (In the experiment, the questions were in German; the screenshots have been translated for this paper.)

Procedure

Each of the participants took part in the following sessions:

- Session 0: Demographic data / pre-test for sampling
- Session 1: Initial training session
- Session 2: Second training session
- Session 3: Evaluation session

Session 2 was scheduled one day after session 1; session 3 took part exactly seven days after session 1. The procedure for each of the sessions was executed as follows.

Session 0: Demographic data / pre-test for sampling

The goal of the first session was to collect demographic data and to do pre-tests to sample three groups that are consistent in terms of their preconditions (ability to concentrate and fine-motor skills). When the participants arrived, we first explained the overall purpose of the study and emphasized that the study was not about evaluating their individual performance, but rather to obtain knowledge about the training process under different conditions in general. We informed the participants that they were taking part on a voluntary basis and that they could exit the study at any point without any personal consequences. We then handed out a questionnaire which asked for various demographic data (e.g. age, gender, handedness) and previous related experiences (assembly work, fischertechnik). After filling out the questionnaire, we conducted two tests with the participants that were used for the sampling: the d2-Revision test for measuring attention and concentration [5] (variable *concentration_performance*) and the 16-step industrial task of the General Assembly Task (GAT) [18] for measuring fine-motor skills (variable *motor_performance* = 300 sec. - task completion time in sec.).

In the first set of participants, all participants finished session 0 before the first participant started session 1, so we were able to balance the participants in the first set as best as possible based on the variables *concentration_performance* and *motor_performance*. Furthermore, the non-native German speakers were distributed among the three groups in order to balance out effects of potential linguistic misunderstandings. In the second set, the participants were added to the group following a greedy strategy: each participant was added to the group in a way so that the variance of the mentioned sampling variables between the groups was minimized.

Session 1: Initial training session

The goal of the second session was to learn the assembly task by heart under one of the three conditions. Participants were instructed to follow the manual or instructions closely and to try to remember each single assembly step as well as the order of steps. In the AR condition, a short explanation about the features of the system was included. For all conditions, participants were free to choose the level of support: in the paper manual condition, people were told to put the manual aside if they did not need it anymore and to pick it up again

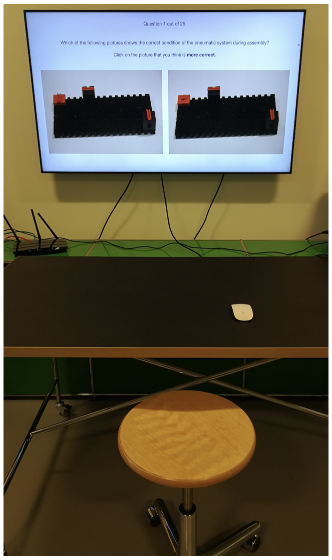

Figure 7. Setup for the quiz during the evaluation session.

if they needed it. In the personal instruction condition, participants were asked to give feedback to the instructor whether they needed more or less information and in the AR condition, participants could control the system via different assistance levels (see section 'Apparatus'). The goal was clearly stated to the participants: "The goal of today's assembly task is to learn to master the assembly process by heart." It was further emphasized that participants had to complete each assembly cycle from the first to the last step before starting over again. A permission was obtained to record the complete training session on video for documentation and evaluation purposes. During the training process, the number of cycles, the time for each of the cycles, the errors and other observations were documented. The training phases were stopped when the participants knew the process by heart, which was considered to be the case under the following circumstances: the participants in the personal training condition completed the assembly without any help or feedback of the instructor; the participants in the AR condition completed the assembly in level 4 (see section 'Apparatus') without any error feedback given by the AR assistive system; the participants in the paper manual condition completed the assembly without looking into the paper manual and were on request absolutely sure to master the process. (Please note that the condition of the paper manual could result in cases where participants learned something systematically wrong, which is a realistic case when someone learns without any external feedback. This happened in two cases; the particular data was excluded from the analysis.) The session was scheduled for 90 minutes, which, for most participants, was more time than they required. However, one participant did not manage to learn it within the given time, so that we excluded the specific data set from the analysis.

Session 2: Second training session

This session took part exactly one day after the previous session and was included in the study to make sure that all participants would reach the same level of knowledge in the training phase. First, the participants were asked to assemble the construction kit one time without any help. The participants were observed and received feedback whether the assembly process and end results were correct or not. In case of errors, the participants were told the exact errors. The participants that made at least one error were asked to deepen their knowledge about the assembly process by another training session, which was conducted in the same way as the day before. Again, the training phase was stopped, when the participants knew the process by heart. The session was scheduled for 45 minutes, even if it took less time for most of the participants. It was recorded on video with the permissions of the participants.

Session 3: Evaluation session

The evaluation session was conducted seven days after the first training session. The aim of this session was to measure the knowledge about the assembly process that the participants still had in mind at that time. For this purpose, two tests were conducted: first, the participants were asked to assemble the construction kit without any help. The participants did not receive any feedback about the correctness of the assembly process. Second, the participants were asked to complete the multiple-choice quiz on a computer screen (see section 'Apparatus'). The setup is shown in Figure 7. The participants answered the 25 questions and the answers were automatically logged by the application.

After finishing the quiz, the participants were thanked for taking part in the study. The participants were asked for a short feedback on their training experience and they had the chance to ask further questions about the study.

RESULTS

In this section, we present the results of the experiment. The data of three participants were removed from the data set for the following reasons: one participant did not complete the training in the first session within the given time frame (AR condition); two participants under the paper condition misinterpreted the manual and executed the task systematically wrong (one of them used wrong pipes, the other participant made systematic orientation errors regarding various parts). We will present the different variables grouped according to the number of training cycles, training time, performance after 24 hours, and performance after one week.

Number of Training Cycles

A Kolmogorov-Smirnov test on the variable n_{day_1} showed that we can assume a normal distribution for n_{day_1} (p = .464). The comparison of the number of training cycles for the three conditions revealed only small differences on the first day (see Figure 8): Participants learning with a paper manual required on average 6.00 cycles (SD = 1.15), followed by participants with a personal training with 6.50 cycles (SD = 1.41) and participants under the AR condition with 7.44 cycles (SD = 1.81). A one-way ANOVA showed no significant difference between the three conditions ($F_{2,21}$ = 1.90, p = .174).

On the second day, however, the differences between the conditions were clearer. Five out of eight participants who learned with a personal training did not make any mistakes and therefore did not have to do any further training sessions, leading

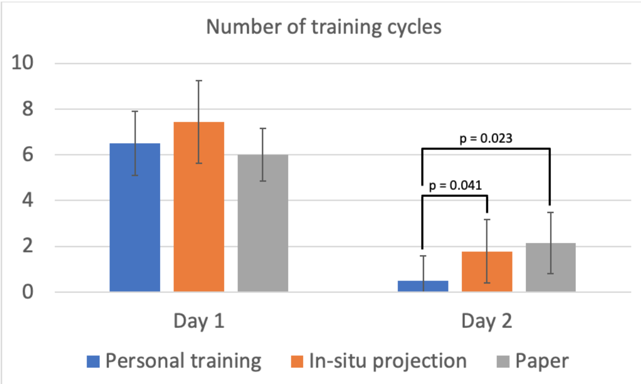

Figure 8. Number of training cycles in each of the two training session.

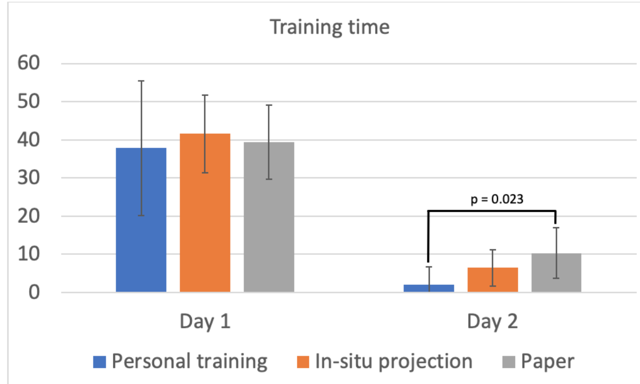

Figure 9. Required training time in minutes for each of the two training session.

to a low average of 0.50 cycles (*SD* = 1.07) in this group. In contrast, participants under the AR conditions had to execute on average 1.78 cycles (*SD* = 1.39) and participants under the paper condition 2.14 cycles (*SD* = 1.35). The Kruskal-Wallis test showed a significant difference between the conditions ($\chi^2 = 6.22$, $p = .045$, $df = 2$). A pairwise comparison using the Mann-Whitney U test showed significant differences between the personal training and the paper condition ($U = 9.5$, $p = .023$) and between the personal training and the AR condition ($U = 16.0$, $p = .041$).

Training Time

The Kolmogorov-Smirnov test on the variables t_{day_1} and t_{day_2} showed that we can assume a normal distribution of the two variables ($p = .555$ for t_{day_1} and $p = .186$ for t_{day_2}).

Corresponding to the number of training cycles, the differences in the training time were small (see Figure 9). The participants in the personal training condition were the fastest with an average training time of 37.80 minutes (*SD* = 17.58 mins.) followed by the paper condition with 39.45 minutes (*SD* = 9.72 mins.) and the AR condition with 41.60 minutes (*SD* = 10.16 mins.). Analyzing the differences with a one-way ANOVA showed no significant differences between the three conditions ($F_{2,21} = 0.18$, $p = .834$).

On the second day, the participants of the personal training group had the shortest average training time with 2.09 minutes (*SD* = 4.66 mins.), followed by the AR group with 6.42 minutes (*SD* = 4.71 mins.) and the paper group with 10.33 minutes (*SD* = 6.55 mins.). Corresponding to the number of training cycles, the one-way ANOVA showed significant differences between the conditions ($F_{2,21} = 4.56$, $p = .023$). Post-hoc comparison using the Scheffé test showed differences between the personal training condition and the paper condition ($p = .023$), but neither between the personal training condition and the AR condition ($p = .263$) nor between the AR condition and the paper condition ($p = .359$).

Performance after One Day

Regarding the errors after 24 hours ($errors\_all_{after\_24h}$), participants made the fewest errors (*M* = 0.88, *SD* = 1.25) in the personal training condition, followed by the AR condition (*M* = 4.44, *SD* = 3.84) and the paper condition (*M* = 5.57, *SD* = 3.78). The results are shown in Figure 10. A Kurskal-Wallis test showed that the differences are significant for the three conditions ($\chi^2 = 8.43$, $p = .015$, $df = 2$). A pairwise comparison using the Mann-Whitney U test showed significant differences between the personal training and paper condition ($U = 5.0$, $p = .007$) and between the personal training and AR condition ($U = 13.0$, $p = .023$), but not between the paper and the AR condition ($U = 27.0$, $p = .632$).

Looking at final errors only ($errors\_final_{after\_24h}$), it was observed that none of the participants under the personal training condition made any final error (*M* = 0.0, *SD* = 0.0), while participants in the paper group made on average 1.29 final errors (*SD* = 1.38) and participants in the AR group 1.44 errors (*SD* = 1.67). Applying the statistic tests as above, again results in significant differences between the personal training and paper condition ($U = 12.00$, $p = .017$) and between the personal training and AR condition ($U = 16.00$, $p = .017$).

Performance after One Week

When measuring the errors after one week ($errors\_all_{after\_one\_week}$), the group under AR condition made the fewest errors (*M* = 2.0, *SD* = 1.87), followed by the paper group (*M* = 2.43, *SD* = 2.51) and the personal training group (*M* = 2.75, *SD* = 3.65). The results are shown in Figure 10. A Kruskal-Wallis test showed no significant differences for the three conditions ($\chi^2 = 0.07$, $p = .996$, $df = 2$). If only final errors are considered, a similar picture emerges: the fewest final errors occured under the paper condition (*M* = 0.29, *SD* = 0.49), followed by the AR condition (*M* = 0.56, *SD* = 1.33) and the personal training condition (*M* = 1.25, *SD* = 2.12). A Kruskal-Wallis test shows no significant differences between the final errors ($\chi^2 = 0.84$, $p = .66$, $df = 2$).

The quiz contained 25 questions in total. For the analysis, we removed one question from the set because we had received feedback from multiple participants that this particular question was misleading and a majority of the participants did not answer this question in a correct way. Out of the 24 questions, participants in the AR group had answered on average 21.00 questions correctly (*SD* = 1.66), followed by the paper

Figure 10. Number of errors in the assembly process.

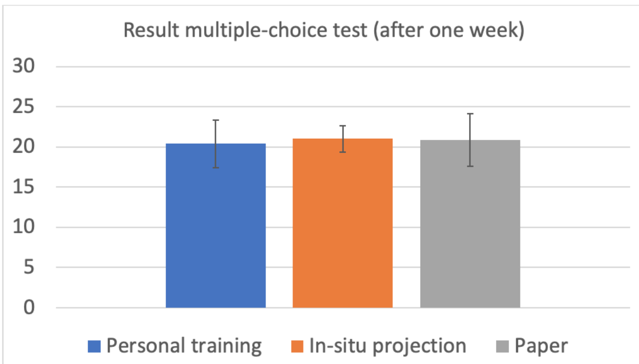

Figure 11. Number of correctly answered questions in the multiple-choice quiz (24 questions in total).

condition (M = 20.86, SD = 3.24) and the personal training condition (M = 20.38, SD = 2.97). The results are shown in Figure 11. The Kruskal-Wallis test shows no significant differences between these final errors (χ^2 = 0.25, p = .88, df = 2).

Finally, we broke down the errors as well as the quiz questions according to specific categories. Again, there are no significant differences between the categories (see Figure 12).

Influence of Other Variables

As discussed, we balanced the variables *concentration_performance* and *motor_performance* among the groups. Here, we analyze the influence of these variables on the results. According to the Kolmogorov-Smirnov test, we can assume normal distribution for both variables (p = .82 for *concentration_performance* and p = .76 for *motor_performance*).

We measured the following *concentration_performance* for the groups: paper condition with M = 153.14 and SD = 34.71, AR condition with M = 172.44 and SD = 32.13 and personal training with M = 150.50 and SD = 41.76. A one-way ANOVA does not show significant differences between the groups ($F_{2,21}$ = 0.92, p = .41). We did not find a correlation between the variables *concentration_performance* and n_{day_1} (r = -.24, n = 24, p = .26), however, we did find a negative correlation

		Personal		In-situ projection		Paper		Kruskal-Wallis	
		M	*SD*	*M*	*SD*	*M*	*SD*	$\chi 2$	*p*
error_all$_{7d}$		2.75	3.65	2.00	1.87	2.43	2.51	0.07	0.97
thereof:	corrected errors	0.63	0.92	0.56	0.73	0.71	0.76	0.24	0.89
	sequence errors	0.88	1.13	0.89	1.17	1.43	1.9	0.3	0.86
	final errors (error_final$_{7d}$)	1.25	2.12	0.56	1.33	0.29	0.49	0.84	0.66
quiz_results$_{7d}$ (total: 24)		20.38	2.97	21.00	1.66	20.86	3.24	0.25	0.88
thereof:	orientation questions (total: 6)	5.25	0.71	5.78	0.44	5.57	0.79	3.02	0.22
	position questions (total: 4)	3.75	0.46	3.78	0.44	3.86	0.38	0.27	0.88
	missing part questions (total: 2)	1.5	0.53	1.56	0.88	1.71	0.49	0.79	0.67
	sequence questions (total: 12)	9.88	2.3	9.89	1.05	9.71	1.8	0.36	0.84

Figure 12. Detailed overview of data from the evaluation broken down into error categories and question categories in the quiz.

between the variables *concentration_performance* and t_{day_1} (r = -.54, n = 24, p < .01). For the other variables there were no correlations with *concentration_performance*.

For *motor_performance* the groups were composed as follows: paper condition with M = 144.71 and SD = 18.74, AR condition with M = 125.67 and SD = 42.00 and personal training with M = 138.88 and SD = 42.00. A one-way ANOVA does not show significant differences between the groups ($F_{2,21}$ = 0.57, p = .57). Analog to *concentration_performance*, there is a negative correlation between the variables *motor_performance* and t_{day_1} (r = -.43, n = 24, p = .04), but not with n_{day_1} (r = -.35, n = 24, p = .10) and neither with any other variable.

We found no correlations between gender and the dependent variables; nor did we find any correlations between age and the dependent variables (the latter might be due to our age-heterogeneous participant group).

DISCUSSION

We will discuss the results according to the hypotheses about training efficiency and knowledge sustainability.

Training Efficiency

With regard to the number of training cycles or the training time, it can be summarized that there were no significant differences on the first day. However, our results of the second day indicate that the participants in the personal training group received a more comprehensive and confident understanding of the assembly process on the first day compared to the other groups (see section 'Performance after one day'). Even though the termination conditions for the training process on the first day were the same for all three groups and the training took about the same time in all groups in the first training session, the data shows clear differences between the participants after 24 hours. Participants from the personal training group were able to remember much better and reproduce the assembly process with fewer errors than participants of the other groups. Five out of eight participants in the personal training group did not produce any errors, while no participant under the paper condition and only one participant under the AR condition could reproduce the task without errors. As a consequence, the training on the second day contained fewer cycles and was shorter for the personal training group compared to the other conditions, so H1 and H3 are supported by our experiment.

Comparing the performance of the paper manual and the AR condition, there is no significant difference, so H2 is not supported in general. However, we have to point out that we had to exclude two participants out of the paper manual condition from the data set who had learned the task systematically wrong. Of course, these particular participants had a high number of training cycles and a long training time on the second day to correct the incorrect understanding of the assembly process. Including these participants into the analysis would have lowered the training efficiency accordingly. So while the projection-based AR system could not outperform the paper manual in terms of training efficiency (which is in congruence with previous performance analysis of the permanent-use scenario, e.g. [6] and [19]), they reliably prevent systematic mislearning through immediate feedback.

Knowledge Sustainability

Looking at the data of the evaluation session one week after the first training session, it can be said that there are no observable differences between any of the three groups, neither in the replication of the task nor in the quiz. Even when we broke down the errors or the questions of the quiz into different categories, there were no significant differences between the groups. This is interesting in so far as the data collected after 24 hours shows a significantly better performance of the participants under the personal training condition. One reason for this difference could be the integration of the second training session into our study setup that potentially balanced out the groups in the following way: participants who remembered the task not that well on the second training day had the possibility to repeat and learn the task again in one to multiple cycles. However, participants who were very confident on the second training day and did not make any errors in the evaluation after 24 hours were not included in the second training, since they fulfilled the termination condition ('no errors in assembly') from the beginning already. Considering the total-time hypothesis [10], the additional training of the less confident participants could potentially balance out the differences on the second training day in the evaluation session.

So, while we observe differences in the training efficiency, we conclude from this observation that, once the knowledge of the assembly process has been built up completely, it continues to exist regardless of the used training method. Therefore, we reject H4 and H5 and accept H6.

LIMITATIONS

Of course, our findings are subject to certain limitations: First, the study results are highly influenced by the specific design and implementations of the instruction types (the particular assistive system and paper manual and the way the trainer acted in the study). While we aimed to keep the information consistent throughout the three conditions (e.g. by using photos in the paper manual and by using identical numbering and colour schemes), obviously this is not possible for all information due to the different nature of the conditions (e.g. in the paper manual there is no way to integrate videos or animated content). The implementation of the assistive system itself affects the findings, e.g. in our study, videos were shown in an endless loop. Multiple participants commented that they wished they had had a feature for restarting, fast forwarding or rewinding the video. Additional features of the system might have a positive impact on the results of the AR condition.

Second, our study was done in a laboratory setup. This includes an artificial task as a replacement for a 'real' industrial task, an artificial environment and a user group that might be different from the real industrial users in terms of their demographics and pre-conditions. The recruitment itself might have an influence as well, as people were recruited through different means (volunteers in first set vs. paid participants in second set) that might influence their engagement in the study.

Third, the sample size of 27 participants who completed the study might be too small to reveal small differences between the conditions. Some variables of our data (e.g. $errors\_all*$ and t_{day_2}) have a high variance, which would require a large sample to discover differences between the groups.

CONCLUSION AND FUTURE WORK

In this paper, we presented the results of an experiment that investigated the use of an projection-based AR assistive system in a training scenario. We compared the training of an assembly task under three different conditions: the AR assistive system, a paper manual and personal training.

The results of our study show that, as expected, the personal training has the best training efficiency. Using an AR assistive system successfully prevents a mislearning of content (as a trainer does), but does not reach the personal training in terms of speed. In congruence with previous studies about projection-based AR systems in permanent-use scenarios, we could not show a higher training efficiency with AR compared to an individual training with a paper manual. Comparing the knowledge sustainability, our study shows that once an assembly task is properly trained, there are no differences in the long-term recall precision, regardless of the training method. Measuring the knowledge of the participants about the assembly process after one week does not show any differences between the groups.

Given these results, we can conclude that AR assistive systems can be used as trainer-less tools for training workers for an assembly task. However, it has to be stated that the benefits of using such a system compared to the low-cost and low-tech variant of a paper manual are – based on our data – restricted to the prevention of mislearning, which of course is an important aspect for quality assurance in industrial production. Further benefits such as a higher training efficiency could not be proven in our study.

Future work should investigate more precisely how certain aspects of a personal training could be integrated into AR assistive systems to improve the training efficiency. By using the latest artificial intelligence (AI) methods and algorithms, the quality of an AR training might be able to move more towards a personal training. Designing and implementing such intelligent assistive systems will be the key challenge for the future training and support of industrial assembly work.

ACKNOWLEDGMENTS

The authors acknowledge the financial support by the Federal Ministry of Education and Research of Germany for the project "Augmented-Reality-Assistenzsysteme für mobile Anwendungsszenarien in der Industrie (FKZ: 03FH005IX6)".

REFERENCES

[1] Mario Aehnelt and Karoline Wegner. 2015. Learn but work!: towards self-directed learning at mobile assembly workplaces. In *Proceedings of the 15th International Conference on Knowledge Technologies and Data-driven Business*. ACM, 17.

[2] Kevin M Baird and Woodrow Barfield. 1999. Evaluating the effectiveness of augmented reality displays for a manual assembly task. *Virtual Reality* 4, 4 (1999), 250–259.

[3] Mark Billinghurst, Mika Hakkarainen, and Charles Woodward. 2008. Augmented Assembly Using a Mobile Phone. In *Proceedings of the 7th International Conference on Mobile and Ubiquitous Multimedia (MUM '08)*. ACM, New York, NY, USA, 84–87. DOI: http://dx.doi.org/10.1145/1543137.1543153

[4] Andrew C Boud, David J Haniff, Chris Baber, and SJ Steiner. 1999. Virtual reality and augmented reality as a training tool for assembly tasks. In *1999 IEEE International Conference on Information Visualization (Cat. No. PR00210)*. IEEE, 32–36.

[5] Rolf Brickenkamp, Lothar Schmidt-Atzert, and Detlev Liepmann. 2010. *Test d2-Revision: Aufmerksamkeits-und Konzentrationstest [Test d2-Revision: Attention and Concentration Test]*. Hogrefe Göttingen.

[6] Sebastian Büttner, Markus Funk, Oliver Sand, and Carsten Röcker. 2016. 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 (PETRA '16)*. ACM, New York, NY, USA, Article 44, 8 pages. DOI: http://dx.doi.org/10.1145/2910674.2910679

[7] Sebastian Büttner, Henrik Mucha, Markus Funk, Thomas Kosch, Mario Aehnelt, Sebastian Robert, and Carsten Röcker. 2017. The Design Space of Augmented and Virtual Reality Applications for Assistive Environments in Manufacturing: A Visual Approach. In *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '17)*. ACM, New York, NY, USA, 433–440. DOI: http://dx.doi.org/10.1145/3056540.3076193

[8] Sebastian Büttner, Oliver Sand, and Carsten Röcker. 2015. 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 (MobileHCI '15)*. ACM, New York, NY, USA, 1130–1133. DOI: http://dx.doi.org/10.1145/2786567.2794342

[9] Sebastian Büttner, Oliver Sand, and Carsten Röcker. 2017. Exploring design opportunities for intelligent worker assistance: a new approach using projetion-based AR and a novel hand-tracking algorithm. In *European Conference on Ambient Intelligence*. Springer, 33–45.

[10] Elaine H Cooper and Allan J Pantle. 1967. The total-time hypothesis in verbal learning. *Psychological bulletin* 68, 4 (1967), 221.

[11] Arindam Dey, Mark Billinghurst, Robert W. Lindeman, and J. Edward Swan. 2018. A Systematic Review of 10 Years of Augmented Reality Usability Studies: 2005 to 2014. *Frontiers in Robotics and AI* 5 (2018), 37. DOI: http://dx.doi.org/10.3389/frobt.2018.00037

[12] Mai ElKomy, Yomna Abdelrahman, Markus Funk, Tilman Dingler, Albrecht Schmidt, and Slim Abdennadher. 2017. ABBAS: An Adaptive Bio-sensors Based Assistive System. In *Proceedings of the 2017 CHI Conference Extended Abstracts on Human Factors in Computing Systems (CHI EA '17)*. ACM, New York, NY, USA, 2543–2550. DOI: http://dx.doi.org/10.1145/3027063.3053179

[13] Christian Dominic Fehling, Andreas Müller, and Mario Aehnelt. 2016. Enhancing vocational training with augmented reality. In *Proceedings of the 16th International Conference on Knowledge Technologies and Data-driven Business*.

[14] Michael Fellmann, Sebastian Robert, Sebastian Büttner, Henrik Mucha, and Carsten Röcker. 2017. Towards a framework for assistance systems to support work processes in smart factories. In *International Cross-Domain Conference for Machine Learning and Knowledge Extraction*. Springer, 59–68.

[15] Pierre Fite-Georgel. 2011. Is there a reality in industrial augmented reality?. In *2011 10th ieee international symposium on mixed and augmented reality*. IEEE, 201–210.

[16] Markus Funk, Andreas Bächler, Liane Bächler, Oliver Korn, Christoph Krieger, Thomas Heidenreich, and Albrecht Schmidt. 2015. Comparing Projected In-situ Feedback at the Manual Assembly Workplace with Impaired Workers. In *Proceedings of the 8th ACM International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '15)*. ACM, New York, NY, USA, Article 1, 8 pages. DOI: http://dx.doi.org/10.1145/2769493.2769496

[17] Markus Funk, Andreas Bächler, Liane Bächler, Thomas Kosch, Thomas Heidenreich, and Albrecht Schmidt. 2017. Working with Augmented Reality?: A Long-Term Analysis of In-Situ Instructions at the Assembly Workplace. In *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '17)*. ACM, New York, NY, USA, 222–229. DOI: http://dx.doi.org/10.1145/3056540.3056548

[18] Markus Funk, Thomas Kosch, Scott W Greenwald, and Albrecht Schmidt. 2015a. A benchmark for interactive augmented reality instructions for assembly tasks. In *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia*. ACM, 253–257.

[19] Markus Funk, Thomas Kosch, and Albrecht Schmidt. 2016. Interactive Worker Assistance: Comparing the Effects of In-situ Projection, Head-mounted Displays, Tablet, and Paper Instructions. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing (UbiComp '16)*. ACM, New York, NY, USA, 934–939. DOI: http://dx.doi.org/10.1145/2971648.2971706

[20] Markus Funk, Sven Mayer, Michael Nistor, and Albrecht Schmidt. 2016. 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 (PETRA '16)*. Association for Computing Machinery, New York, NY, USA, Article Article 45, 4 pages. DOI: http://dx.doi.org/10.1145/2910674.2910730

[21] Markus Funk, Sven Mayer, and Albrecht Schmidt. 2015b. Using In-Situ Projection to Support Cognitively Impaired Workers at the Workplace. In *Proceedings of the 17th International ACM SIGACCESS Conference on Computers & Accessibility (ASSETS '15)*. ACM, New York, NY, USA, 185–192. DOI: http://dx.doi.org/10.1145/2700648.2809853

[22] Dominic Gorecky, Ricardo Campos, Harish Chakravarthy, Rüdiger Dabelow, Jochen Schlick, and Detlef Zühlke. 2013. MASTERING MASS CUSTOMIZATION–A CONCEPT FOR ADVANCED, HUMAN-CENTERED ASSEMBLY. *Academic Journal of Manufacturing Engineering* 11, 2 (2013).

[23] Sandra G Hart. 2006. NASA-task load index (NASA-TLX); 20 years later. In *Proceedings of the human factors and ergonomics society annual meeting*, Vol. 50. Sage publications Sage CA: Los Angeles, CA, 904–908.

[24] Bui Minh Khuong, Kiyoshi Kiyokawa, Andrew Miller, Joseph J La Viola, Tomohiro Mashita, and Haruo Takemura. 2014. The effectiveness of an AR-based context-aware assembly support system in object assembly. In *2014 IEEE Virtual Reality (VR)*. IEEE, 57–62.

[25] Oliver Korn, Markus Funk, and Albrecht Schmidt. 2015. 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 (EICS '15)*. ACM, New York, NY, USA, 84–93. DOI: http://dx.doi.org/10.1145/2774225.2774834

[26] Thomas Kosch, Romina Kettner, Markus Funk, and Albrecht Schmidt. 2016. Comparing tactile, auditory, and visual assembly error-feedback for workers with cognitive impairments. In *Proceedings of the 18th International ACM SIGACCESS Conference on Computers and Accessibility*. ACM, 53–60.

[27] Michael R Marner, Andrew Irlitti, and Bruce H Thomas. 2013. Improving procedural task performance with augmented reality annotations. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 39–48.

[28] Bernd Schwald and Blandine De Laval. 2003. An augmented reality system for training and assistance to maintenance in the industrial context. (2003).

[29] Björn Schwerdtfeger, Daniel Pustka, Andreas Hofhauser, and Gudrun Klinker. 2008. Using laser projectors for augmented reality. In *Proceedings of the 2008 ACM symposium on Virtual reality software and technology*. Citeseer, 134–137.

[30] Arthur Tang, Charles Owen, Frank Biocca, and Weimin Mou. 2003. Comparative Effectiveness of Augmented Reality in Object Assembly. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '03)*. ACM, New York, NY, USA, 73–80. DOI: http://dx.doi.org/10.1145/642611.642626

[31] Xiangyu Wang, Soh K Ong, and Andrew YC Nee. 2016. A comprehensive survey of augmented reality assembly research. *Advances in Manufacturing* 4, 1 (2016), 1–22.

[32] Giles Westerfield, Antonija Mitrovic, and Mark Billinghurst. 2013. Intelligent augmented reality training for assembly tasks. In *International Conference on Artificial Intelligence in Education*. Springer, 542–551.

[33] Giles Westerfield, Antonija Mitrovic, and Mark Billinghurst. 2015. Intelligent augmented reality training for motherboard assembly. *International Journal of Artificial Intelligence in Education* 25, 1 (2015), 157–172.

[34] Ellen S Wilschut, Reinier Könemann, Molly S Murphy, Gu JW Van Rhijn, and Tim Bosch. 2019. Evaluating learning approaches for product assembly: using chunking of instructions, spatial augmented reality and display based work instructions. In *Proceedings of the 12th ACM International Conference on PErvasive Technologies Related to Assistive Environments (PETRA '19)*. ACM, 376–381.

[35] Xianjun Sam Zheng, Cedric Foucault, Patrik Matos da Silva, Siddharth Dasari, Tao Yang, and Stuart Goose. 2015. 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 (CHI '15)*. ACM, New York, NY, USA, 2125–2134. DOI: http://dx.doi.org/10.1145/2702123.2702305