Comparison of Different Types of Augmented Reality Visualizations for Instructions

Florian Jasche
florian.jasche@uni-siegen.de
University of Siegen
Siegen, Germany

Sven Hoffmann
sven.hoffmann@uni-siegen.de
University of Siegen
Siegen, Germany

Thomas Ludwig
thomas.ludwig@uni-siegen.de
University of Siegen
Siegen, Germany

Volker Wulf
volker.wulf@uni-siegen.de
University of Siegen
Siegen, Germany

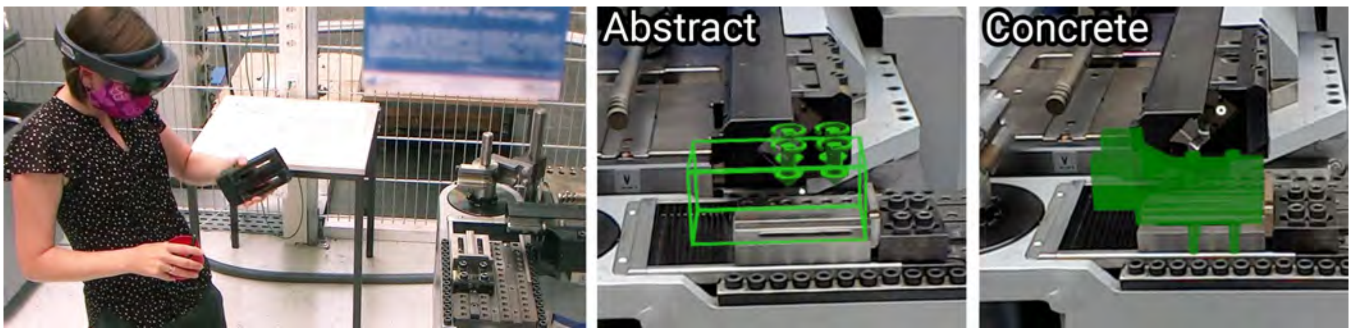

Figure 1: In our study we compare abstract and concrete AR visualizations for instructions using head mounted displays.

ABSTRACT

Augmented Reality (AR) is increasingly being used for providing guidance and supporting troubleshooting in industrial settings. While the general application of AR has been shown to provide clear benefits regarding physical tasks, it is important to understand how different visualization types influence user's performance during the execution of the tasks. Previous studies evaluating AR and user's performance compared different media types or types of AR hardware as opposed to different types of visualization for the same hardware type. This paper provides details of our comparative study in which we identified the influence of visualization types on the performance of complex machine set-up processes. Although our results show clear advantages to using concrete rather than abstract visualizations, we also find abstract visualizations coupled with videos leads to similar user performance as with concrete visualizations.

CHI '21, May 8–13, 2021, Yokohama, Japan

ACM ISBN 978-1-4503-8096-6/21/05…$15.00
https://doi.org/10.1145/3411764.3445724

CCS CONCEPTS

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

KEYWORDS

Augmented Reality, Visualization, User Study, Instructions

ACM Reference Format:

Florian Jasche, Sven Hoffmann, Thomas Ludwig, and Volker Wulf. 2021. Comparison of Different Types of Augmented Reality Visualizations for Instructions. In *CHI Conference on Human Factors in Computing Systems (CHI '21), May 8–13, 2021, Yokohama, Japan.* ACM, New York, NY, USA, 13 pages. https://doi.org/10.1145/3411764.3445724

1 INTRODUCTION

In our daily work we often find ourselves in situations in which we encounter problems. If we are unable to solve the problems ourselves, we may seek support from people with more experience and expertise [10, 30]. This is particularly true within complex work environments and when performing knowledge-intensive tasks. These tasks encompass non-trivial activities that have a strong contextual reference characterized by a multitude of environmental factors, requiring a significant amount of knowledge to achieve a successful outcome. Examples of knowledge-intensive tasks are setting up or retooling a modern production machine, repairing a car engine, or solving 3D printer errors.

In recent years, the use of Augmented Reality (AR)-based technologies to support remote troubleshooting [3, 25, 28] as well as digitally enriched instructions [8, 27] has increased significantly for these types of tasks. Moreover, the COVID-19 pandemic and the resulting lock-down in many regions of the world have created increased interest in AR for supporting remote maintenance or troubleshooting and detailed instructions [44]. The primary advantage of AR is that it creates a 3D-based overlay of digital content onto the physical environment, which offers a rich set of additional information [1, 13, 23]. While the general application of AR concepts shows clear benefits for supporting complex settings, its actual design and technology as well as its impacts on users are often not well established. Regardless of the chosen hardware and its advantages in specific applications (such as head-mounted displays [HMDs] for hands-free applications [29, 32], hand-held displays [HHDs] like smartphones for cost-effective, easy-to-use cases [2, 34], or projector technologies for longer, static, and group settings [26, 33]), a crucial factor is the actual visualization and presentation of AR content in relation to the complex tasks and knowledge-intensive work.

Existing studies evaluating different AR-based content visualizations and their performance on working practices have compared different media types (e.g., paper-based instructions versus AR-based visualization) or different types of AR hardware (e.g., hand-held devices versus head-mounted displays versus projection-based AR). Our study complements this set of existing studies by evaluating different forms of visualization on the same hardware type for performing a complex task. In particular, we examine the task completion time, error rate, and task and cognitive load of setting up a complex machine with abstract and concrete visualization types on the head-mounted display Microsoft HoloLens. The visualization types consist of abstract and rather simple holograms like wire frame boxes and 3D arrows, as well as concrete detailed 3D models from CAD data.

In the following sections, we present the current state of the art regarding the use of AR in manual assembly and maintenance tasks (Section 2). We then identify and discuss the research gap regarding the comparison of different visualization types on the same task and same hardware (Section 3). In Section 4, we describe the applications we compared within our evaluation study. We then present our study design and procedure in Section 5 and our results in Section 6. Finally, we conclude with a discussion of the results with regard to visualization types of AR applications and their implications for AR authoring tools (Section 7).

2 RELATED WORK: AUGMENTED REALITY FOR SUPPORTING COMPLEX TASKS

Augmented Reality has a wide application field, particularly in the industrial context, where there is a huge interest in using AR for providing assistance for complex tasks such as manual assembly or maintenance work [13, 45]. In these applications, virtual computer-generated visual features are superimposed on the real environment support the execution of these complex tasks. These AR applications commonly support workers in sequential tasks by providing context- and task-related information in-situ. Over the last two decades there have been several studies investigating the suitability and effectiveness of AR applications for manual assembly or maintenance tasks. In these studies, AR applications were usually compared to paper-based instructions by measuring the time required to complete a task, the amount of errors that occurred during the task, and the mental workload during the execution.

Funk et al. [16] compared four different hardware types (paper, a smart glass, a smartphone, and in-situ projection) using a standardized Lego Duplo task [15] with 16 participants (with-in subject). In this study, all hardware types used the same pictorial instructions. The authors measured the task completion time, errors, and the task load. The results showed that there is no significant difference between in-situ projections and paper instructions regarding the completion time. However, the same pictorial instruction on a HMD and HHD resulted in a significantly slower time to recognize the correct elements compared to the paper. In addition, more errors were produced when using the HMD compared to the hand-held device and in-situ projection. In regards to the cognitive load, the in-situ projection was perceived to be the lowest and the HMD was the highest.

Blattgerste et al. [6] investigated different in-situ instructions for assembly tasks by comparing four visualization types (3D in-situ, 2D in-situ, 3D wire, and side by side) using the Microsoft HoloLens in a with-in subject study consisting of 24 participants asked to solve a standardized LEGO Duplo task [15]. While the results show a faster completion time and less errors when using 3D in-situ visualizations, there was no significant difference between the systems in terms of the task load. Blattgerste et al. [7] further compared different types of AR hardware (Microsoft HoloLens, Epson Moverio BT-200, and smartphone) to paper-based instructions. The paper-based instructions resulted in the fastest overall task completion time, while the Microsoft HoloLens led to the least number of errors, but with a significantly higher cognitive load.

Smith et al. [37] analyzed the effect of different interaction modalities (touch and voice) and visualization modes (3D model, text annotation, and in-situ video) of an AR application on task load, task completion time, and error rate for 24 participants performing a LEGO Duplo task. The in-situ video led to the fastest task completion time, followed by superimposed 3D models and text annotations. In addition, the task load was the lowest for the in-situ videos and the highest for the text annotations. The method by which the users interact with the AR content (touch versus voice) had little to no effect on the measured performance.

Radkowski et al. [31] investigated the effect of different visual features (abstract versus concrete) of 3D content for different degrees of difficulty in manual assembly tasks. Abstract refers to visual forms (e.g., lines, colors, shapes) that are separate from any concrete 3D models. The authors conducted a between-subject study with 33 participants (abstract visualization versus concrete visualization versus paper-based). The participants had to assemble a mechanical axial piston engine with a total of 16 manual assembly process steps. Two of these steps were rated with a high degree of difficulty; all the others were rated with a low degree. In comparison to the LEGO Duplo tasks [15], this assembly task can be considered to be more difficult as it is based on real conditions. The authors used a tabletop AR workstation displaying an AR view of the working space. They showed that the abstract visualization led to an overall longer completion time and more errors compared to concrete visualizations. Further, the concrete AR visualization resulted on average in a similar completion time compared to the paper-based instructions,

Table 1: Overview of the closely related studies. Most AR vs. AR studies also used paper instructions as baseline.

	Study	Medium	Task	AR-Visualization
Paper vs. AR	Tang et al. 2003 [40]	Paper vs. Sony Glasstron LDI-100B	Lego Duplo	3D models and arrows
	Syberfeldt et al. 2015 [38]	Paper vs. AR-enabled Oculus Rift	3D puzzle	Abstract highlighting
	Fiorentino et al. 2014 [14]	Paper vs. Monitor-based AR	Engine maintenance	3D animations, text, voice
	Hou et al. 2015 [22]	Paper vs. Monitor-based AR	Pipe system assembly	3D models of pipes
	Uva et al. 2018 [42]	Paper vs. In-situ Projection	Engine maintenance	Test and highlighting
AR vs. AR	Funk et al. 2016 [16]	Tablet (no AR) vs. Epson Moverio BT-200 vs. In-Situ Projection	Lego Duplo	Same pictorial instructions
	Blattgerste et al. 2017 [7]	Epson Moverio BT-200 vs. Microsoft HoloLens vs. Smartphone	Lego Duplo	Same, but slightly adjusted for the different media
	Smith et al. 2020 [37]	Smartphone	Lego Duplo	3D models vs. text-annotation vs. in-situ video
	Radkowski et al. 2015 [31]	Monitor-based AR	Engine maintenance	Abstract vs. concrete
	Blattgerste et al. 2018 [6]	Microsoft HoloLens	Lego Duplo	3D in-situ vs. 2D in-situ vs. 3D wire vs. side-by-side

while the completion time for abstract visualization was almost twice as long. This result suggests that concrete visualization is more suitable for relatively simple tasks. However, which kind of visualization type is more appropriate for rather complex tasks could not be answered as there were ambiguous results regarding the time of the two difficult assembly steps.

Wiedenmaier [47] argues that users only benefit from AR when focusing on difficult assembly tasks and most of the previous studies [7, 16, 31, 37] confirmed that AR does not provide an advantage for the completion time of simple tasks. Furthermore, Wiedenmaier claimed that "[g]ood manuals often use line drawings instead of photos, to reduce complexity. In the same way, assembly tasks can often be simplified by more abstract objects"[47]. Haller [18] concluded that 3D models look nice for AR developers, however the actual end users often ask for abstract but familiar schemes or 2D drawings. These statements contradict the results of Blattgerste et al. [6] and Radkowski et al. [31].

Interactivity is considered as an important factor when focusing on AR applications [4], but only few studies (e.g. [37]) have investigated the impact of different interaction techniques in this field. In most existing studies regarding types of instruction systems, there is no active interaction with the system itself. Usually, techniques such as Wizard of Oz or automatic detection of the correct placement are used and the user should fully concentrate on the task. However, an automatic detection of the correct placement can not be implemented for every task, wherefore multi-modal interactions with hand gestures, voice or special controllers are meant to enhance the user experience in AR applications [9, 20]. While designing the interaction, the environment as well as the task needs to be considered. Due to loud noise, for example on the shop floor, voice interaction is often not suitable for interacting with the AR system.

3 RESEARCH GAP AND APPROACH

The literature analysis reveals many existing studies that evaluate AR-based assistance for manual assembly and maintenance tasks, as shown in Table 1. Usually these studies either compare AR-based instructions to those using different media such as paper-based and mentor-based instructions [46], or they focus on comparing AR-based instructions through different hardware types such as projection-based AR, hand-held devices, or HMDs. Only a few studies compare different visualization types for instructions using the same AR hardware type.

Although the success of AR-based applications for assembly depends on many different factors, current studies almost do not adequately address the complexity and setting of the task [11] and instead focus on simplified Lego assembly tasks, neglecting a variety of conditions of real complex tasks. Only one study [31] compares visualizations of one type of hardware in real tasks, but the static, monitor-based apparatus of the study setup and the resulting limited interaction and visualization area neglect the complexity of mobile settings as well as limit the possibilities of 3D representation.

Through our study, we contribute to the field of comparative studies of AR for assembly tasks by examining the visualization types for instructions that are suitable for HMDs in real-world tasks. To accomplish this, we conducted a user study to identify the influence of different abstract and concrete visualization types on the key performance indicators in machine set-up processes.

For our different forms of visualization, we distinguish between the types of 3D holograms. We use very simple (abstract) holograms like wire mesh boxes and simple 3D arrows as well as detailed (concrete) 3D models from CAD data; we refer to the former as abstract AR (AAR) and the latter as concrete AR (CAR). In addition, we also distinguish between the use of multimedia data through videos. Such instructions with additional video files we refer to as abstract AR with videos (AAR+V) and concrete AR with videos (CAR+V). Additionally, we used paper-based instructions as a base line. The key performance indicators of our study were the task completion time, error rate, as well as task and cognitive load. We aim to investigate the following hypotheses:

- H1: The task completion times for the set-up procedures with AAR, CAR, AAR+V, CAR+V and paper-based instructions are significantly different.
- H2: The number of errors for the set-up procedures with AAR, CAR, AAR+V, CAR+V and paper-based instructions are significantly different.
- H3: The cognitive load for the set-up procedures with AAR, CAR, AAR+V, CAR+V and paper-based instructions is significantly different.

To verify our hypotheses, we carried out a between-subject study as the execution of the instruction with either visualization type may lead to learning effects which could influence our dependent variables. We collected qualitative as well as quantitative data sets. With regard to the quantitative data set, we defined dependent variables such as task completion time, error rate, mental workload, perceived usability and user experience and the visualization type with five characteristics as independent variable.

4 AR APPLICATIONS

We evaluated two self-developed Microsoft HoloLens applications that support the same task of performing hardware-centered activities – but with different visualizations. The main concept of both applications is to guide users in performing a task through AR instructions. The applications were meant to support novice users in procedural machine set-up tasks.

When we compare concrete (see Figure 2) with abstract (see Figure 3) types of visualization, it is clear that concrete visualization is much more detailed and closer to the actual application. While in the abstract visualization only arrows or generic forms indicate points of action, the concrete visualization maps exactly the 3D models with the real work pieces, giving the user richer indications of the parts to be used and their locations.

4.1 Concrete Augmented Reality Visualizations

The first application (concrete augmented reality, CAR) was created in the research project Cyberrüsten 4.0 and focuses on the procedure of setting up complex machines [21]. The application uses 3D models provided by CAD machine data and the instructions are typically created by professionals on the shop floor (domain experts) with the same AR application in edit mode. The instructions are organized in small sub-tasks as sequential step-by-step commands.

When creating an instruction with the CAR application the expert defines which component needs attention by selecting the component's name from a list, selecting a virtual overlay (shadow) of the component on the machine, then defining the mounting direction for each step. The expert can take photos or videos to further visually illustrate the activities for each step. Photos and videos are recorded with the built-in RGB camera of the Microsoft HoloLens to create the same perspective as that of the expert.

To view the instructions, a main panel shows the information about the current process step including the current step number and the name of the component. On the left side of the main panel is another panel showing pictures and videos of the current step. By default, the panels are connected to each other and follow the users' movement, but they can be fixed to any position within the room by the user. Buttons are located in the lower corners of the main panel to navigate through the instructions. Furthermore, concrete 3D models show the component and its mounting position, as shown in Figure 2. The 3D models move to the mounting position to show the mounting direction, the models dwell for three seconds on the mounting position, and then the animation repeats.

4.2 Abstract Augmented Reality Visualizations

The first field trial at a local company of our first application showed that creating the AR content and preparing the CAD data required an enormous amount of effort, which may not be possible by actual domain experts. Furthermore, the company in which we evaluated the first application wanted to use it for other tasks (beyond setting up a machine) for which no CAD data were available. We therefore implemented a second AR application (abstract augmented reality, AAR) from scratch with less detailed 3D models compared to the first application which also simplified the content creation. This ensured that the actual AR content creation could be done by non-AR experts. While building this second application, we replaced the detailed CAD models with simple abstract 3D models, such as arrows, circular arrows, and boxes (see Figure 3).

In this second application, the creation of the instructions can be completely done by domain experts. The expert can therefore select different 3D models from a table for every step and place, and rotate and scale them using hand gestures. For each step, a free text description can be added. Similar to the first application, the expert can document each step via pictures or videos with the support of the built-in RGB camera of the Microsoft HoloLens. The interface for viewing the instructions within the AAR application follows the same design as the CAR application. The simple 3D models are shown on the machine, as shown in Figure 3.

5 STUDY

5.1 Setup and Apparatus

We collected the data based on the performance of an identical complex task. We instructed the participants to equip a Wafios RBV 35 bending machine with a set of bending tools using the predefined visualization types (paper, AAR, AAR+V, CAR, CAR+V). At the beginning of the task, no tools were mounted on the machine and the machine axes were set to an initial position, which did

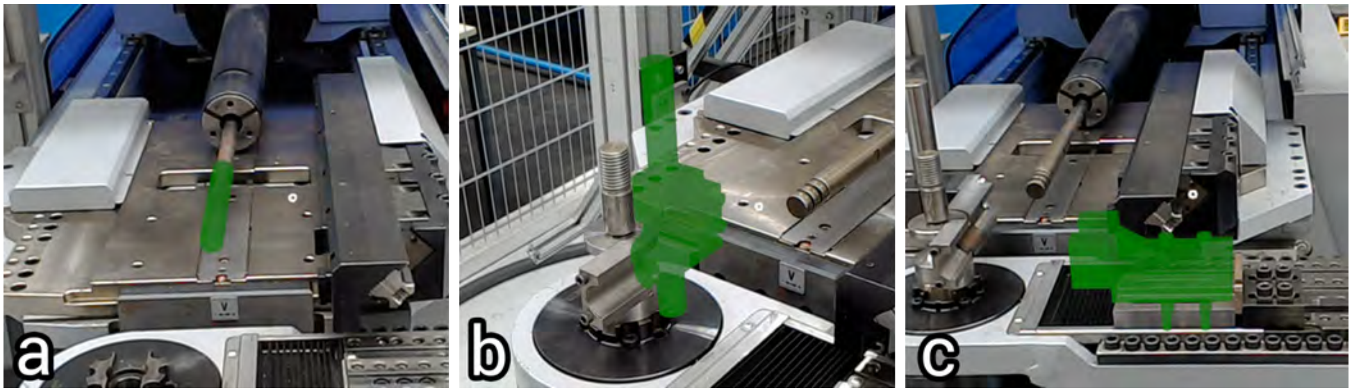

Figure 2: Concrete AR visualization of assembly step a) 1, b) 7, c) 11.

Figure 3: Abstract AR visualization of assembly step a) 1, b) 7, c) 11.

not change during the course of the study. The individual tool components were lined up on a table and individually labeled. The position of the tools was identical during each study.

The participants were given the task to mount the tool components on the machine according to the instructions. The instructions specified a certain sequence which was identical for all visualization types. The participants were instructed to follow the sequence, but they were allowed to go back and forth between the individual steps. In total, the entire process included eleven steps, each with one tool component. Tool component nos. 5, 8 and 11 include additional screws. No other tools (e.g. Allen wrenches) were required for the execution of the steps, as the screws only required hand tightening. This is slightly different in reality, but it makes the entire evaluation setup easier. Figure 5 shows the dependencies of the individual tool components. Some tool components do not have mechanical connections with others. For example, tool component no. 1 could also have been mounted in the last position. In contrast, it was necessary to install tool component no. 2 before no. 3. The situation was made more difficult by the fact that tool no. 4 could only be fitted correctly if nos. 2 and 3 were fitted correctly, see Figure 4. This realistic conception of the task thus included complex steps that are difficult to perform without prior knowledge. We determined the complexity for each step according to various criteria (see for example [31]) on a 4-point scale from low to high complexity.

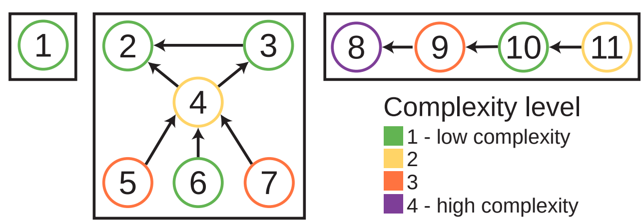

Figure 4: Dependency graph of the assembly steps. Each box indicates an assembly group. There are no dependencies between the assembly groups and the assembly order could be changed. The arrows indicate the dependencies of the individual steps, e.g. step 5 depends on the correct assembly of step 4. The color of the circles indicates the complexity level of the step.

For each step, we created paper instructions and content for the AR applications. We selected and placed holograms for each step and recorded a short video showing the task execution from a first-person perspective. Within the AR applications, the name of the components and the holograms were shown. There were no additional textual instructions. Only for the AAR+V and CAR+V, further videos were provided. In the paper instructions we included a single picture of the mounted component and another picture of the component from the CAD data, as seen in Figure 5. Additionally, the step number, name of the component and a one-sentence description were provided next to the picture. Each step was printed on one page in landscape orientation. All instructions were evaluated and approved by a mechanical engineer with domain knowledge.

Figure 5: Representations of the tool for each step.

5.2 Data collection and processing

We collected several kinds of data during the execution of the study. The execution times of each individual assembly step were recorded by using an integrated logging function of the HoloLens applications. Specifically, the dwell time (including the time for watching the videos) of the participants in each step was logged and added to a total time at the end. The start and end of each step is indicated by pressing the next button in the main panel. For accurate measurements, e.g. to compensate orientation time in the beginning, we added one dummy step before and after the actual task. The entire procedure was also recorded with a video camera. The logged data were subsequently checked against the video and adjusted if the participants accidentally opened the bloom menu and did not know how to close it. In this case, we intervened and later removed this period from the total completion time. The errors were recorded manually and documented with pictures after the completion of the procedure. An error was counted when the position or orientation did not match the specifications of an instruction step. We did not distinguish the error type; however, subsequent errors were not counted if the relation to the dependent components was not broken. In total eleven errors are possible, one for each step.

We further collected additional data using four standardized questionnaires. We used the (1) Rating Scale Mental Effort (RSME) [49] and the (2) NASA-TLX [19] questionnaires for measuring the task load, the (3) System Usability Scale (SUS) [5] for measuring the perceived usability of the system, and the (4) User Experience Questionnaire (UEQ) [24] for measuring the user experience.

To enrich our quantitative data, we additionally video-recorded the execution of the task and audio-recorded a follow-up discussion which was carried out as a closing interview. This discussion followed a predefined interview guideline. The audio recordings were fully transcribed and coded using a deductive coding system which was derived from the interview guideline.

5.3 Participants

We recruited 48 voluntary participants in total (7 female and 41 male) through social media and public mailing lists. Most of them (33) are undergraduate students of different disciplines (21 industrial engineering, 5 human-computer interaction, 6 mechanical engineering, 1 business administration). The remaining 15 participants are professionals in other industries (e.g., banker, journalist, soldier and software developer). We ensured that none of the participants had experience in operating the machine. Five persons had previous experience with setting up machines. We divided the participants into five groups by separating them equally according to their jobs and prior knowledge in mechanical tasks and setting up machines. From the participants in the AR conditions, 15 participants did not know about AR, 18 participants knew the concept of AR but had never used an AR application, and 7 participants had previously used some kind of AR (3 of these participants had previously used the Microsoft HoloLens).

5.4 Procedure

We began our study by informing each participant about the overall goal and the procedure. All participants signed a consent at the beginning of the test session informing them about their right to stop participating in the study at any time without any consequences. Furthermore, we informed the participants that the session would be video- and audio-recorded, but all recordings would be treated confidentially, and references would only be made in anonymized form. The participants had sufficient time to read the consent form and ask questions before signing the form.

Before we began with the actual task, we first calibrated the Microsoft HoloLens for each participant with the built-in calibration application. Next, we started one of our two applications and used a sample instruction for explaining the application and the interaction with the Microsoft HoloLens. The process lasted until we were sure that the users had mastered the most important gestures, especially the AirTap. Once this was achieved, we moved to the actual machine and we briefly explained the general task. We also showed the participants where the components were located and told them that they were completely on their own and would not receive any support. After the task was completed, we removed the Microsoft HoloLens, cleaned it for the next participant, and asked the participant to answer the RSME and NASA-TLX questionnaires. We then discussed the task with the participants near the machine and asked them to comment on the individual steps and rate them according to the perceived complexity. For this purpose, a 4-point scale was taken, where 1 corresponds to low perceived complexity and 4 to high perceived complexity. The participants were then asked to complete the UEQ and SUS questionnaires. Finally, the procedure was completed with a closing semi-structured interview. Overall, the procedure lasted an average of 70 minutes.

Of note, due to the ongoing COVID19-pandemic, we and the participants wore oronasal masks and maintained a minimum 1.5-meter (around 5 ft.) distance. In addition, the test area and gear were cleaned and disinfected vigorously before and after every session.

6 RESULTS

With the data collected from the participants, we statistically analyzed the values of the dependent variables. As the one-way ANOVA requires normal distribution and homogeneity of variance, we first tested for normal distribution using the Shapiro-Wilk-Test [36] and for the homogeneity of variance using the Levene-Test [41]. The assumption of normal distribution was invalidated by the error rate, and homoscedasticity was violated by the RSME values. For the variables with normal distribution and homoscedasticity we used a one-way ANOVA for the analysis, the error rate was analyzed with a Kruskal-Wallis test [43] and the RSME values with a Welch-ANOVA [41]. We further used the Bonferroni post hoc test [39] for the one-way ANOVA and Games-Howell post hoc test for the Welch-ANOVA.

Table 2: Test results for normal distribution and homoscedasticity.

Dependent Variable	Shapiro-Wilk-Test		Levene-Test	
	Sig. (p)	Result	Sig. (p)	Result
Task Completion Time	.102	OK	.499	OK
Error Rate	.008	NO	.293	OK
RSME	.461	OK	.001	NO
NASA-TLX	.631	OK	.140	OK

6.1 Task Completion Time

The first hypothesis (H1), which states that the mean values of task completion time differ significantly between the different visualization types, is not confirmed ($F(4, 43) = .194$, $p < .940$) by the one-way ANOVA. A closer look at the Bonferroni post-hoc test consequently shows that no combination of the dependent variables reveals a significant difference between the mean values of the execution times. The descriptive statistics show that the paper instruction required the least mean time for completing the whole set-up procedure with 646.5s (SD = 237.8s), followed by AAR with 675.0s (SD = 171.6s), and AAR+V with 677.9s (SD = 149.6s). The concrete visualization types both required more time, with 708.0s for CAR (SD = 206.1s) and 712.0s for CAR+V (SD = 152.9s).

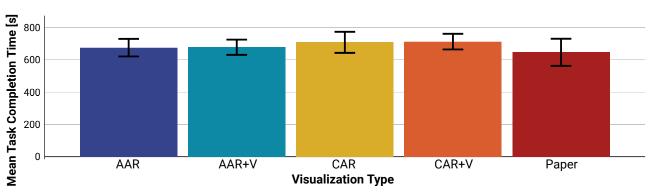

Figure 6: The mean task completion time using the different visualization types. Error bars represent the standard error of the mean.

There was a huge difference in the usage of the videos. According to the HoloLens log, the participants using AAR+V watched the videos more frequently than those using CAR+V. On average, the participants using AAR+V watched the videos 18.5 times (SD = 2.77) whereas participants using CAR+V only 11.3 times (SD = 6.59). Participant CAR+V-8 did not even watch a video once. CAR+V-2 stated: "*I watched the videos for details. Sometimes you did not even need the video because the task was so simple.*" The time for watching the videos is included in the total completion time. However, there was no correlation between the number of videos watched and the task completion time.

6.2 Errors

The second hypothesis (H2) states that the mean number of errors differs significantly between the different visualization types. In this case the number of errors does not show a normal distribution; therefore, we performed the Kruskal-Wallis test. The results show that the mean number of errors differs significantly (Chi-square(2) = 21.671, $p < .001$). Descriptive statistics show that participants with a CAR+V visualization made only 1.7 errors on average (SD = 0.67). The average number of errors for paper and AAR+V was 2 (SD = 1.19 and SD = 0.66). Participants with AR visualizations without video had the highest number of errors. The average number of errors for CAR was 2.7 (SD = 0.82) and for AAR was 3.9 (SD = 1.1).

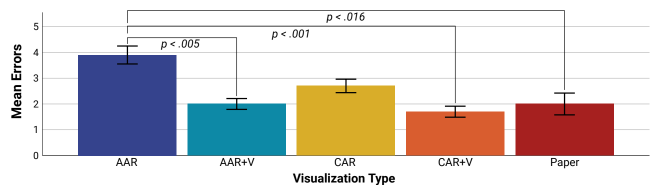

Figure 7: The mean number of errors that were made during the task using the different visualization types. Error bars represent the standard error of the mean. The significance lines show the significance value of the pairwise comparison of the Krukall-Wallis Test.

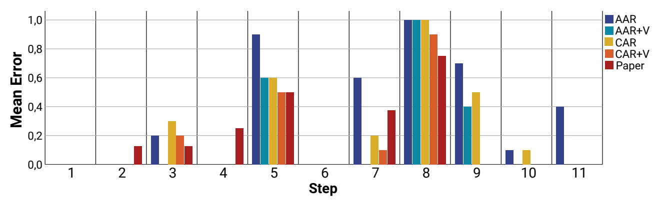

Figure 8: Distribution of mean errors by step, grouped by visualization type.

A closer look at the pairwise comparison shows that CAR+V and AAR differ significantly ($p < .001$) with a large effect size ($r = .912$). On average, participants that used CAR+V made 2.2 fewer errors than those using AAR. The same effect holds for AAR and AAR+V. Participants using AAR+V had an average of 1.9 fewer errors ($p < .005$) with a large effect size ($r = .774$). Even participants with paper instructions, with 1.9 fewer errors on average, had a lower number of errors and are therefore significantly ($p < .016$) different from the AAR instruction with a large effect size ($r = .742$). The results show that participants using abstract AR visualizations without video perform poorly in this setting, as they made a particularly high number of errors. The average number of errors associated with AAR+V is significantly lower compared to AAR.

The differentiated consideration of the error mean values related to the individual steps and the visualization types is shown in Figure 8. We clearly saw that, on average, most errors occur in set-up steps 5 and 8. The number of errors reflect the pre-determined complexity of the individual steps. However, the perceived complexity was not in line with the pre-determined complexity, except step 1 and 6.

Overall, the participants rated low complexity steps higher and high complexity steps lower. For example, step 8 was rated with an average of 2 by the participants versus our pre-determined complexity of 4. Furthermore, there is no tendency that indicates that videos lead to a overall lower perceived complexity.

6.3 Mental Effort

The third hypothesis (H3) states that the mean values of the indicators for cognitive load differ significantly between the different visualization types. For this purpose, we investigated the results of the RMSE and NASA-TLX questionnaires. The RSME values did not show variance homogeneity and therefore the Welch-ANOVA was investigated. The results show that there is a significant difference between the groups ($F(4, 21.325) = 3.080$, $p < .038$). The Games-Howell post-hoc test does not show significant differences between the groups. Only between AAR and paper was the significance level almost reached with $p < .055$. Alternatively, the Bonferroni Post-hoc test was applied, which shows a significant difference between AAR and Paper ($p > .010$) with a medium effect size ($f = .26$).

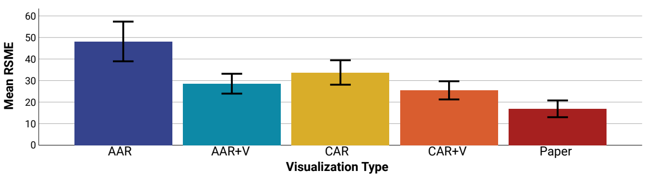

Figure 9: The mean mental effort during the task using the different visualization types. Error bars represent the standard error of the mean.

An examination of the descriptive statistics reveals that the RSME mean for using paper at 16.87 (SD = 10.99) is much lower compared to AAR at 48.15 (SD = 29.15). Both AAR and CAR result in a higher RSME value compared to paper instructions. When comparing CAR with 33.75 (SD = 17.91) and AAR, CAR tends to be slightly lower. The same applies to AAR+V with 28.55 (SD = 14.57) and CAR+V with 25.45 (SD = 13.33). These results show that videos lead to lower RSME values, which indicates less stress compared to set-ups without video.

6.4 Task Load

The NASA-TLX values show variance homogeneity and were tested for significance using one-way ANOVA. The test shows significant differences between the groups ($F(4, 43) = 2{,}695$, $p < .043$). In contrast, the Bonferroni post-hoc test shows no significant difference between the groups. Only CAR and CAR+V are approximately significantly different ($p < .057$) with a small effect size ($f = .204$).

The descriptive statistics provide a good indication that the mean value of CAR with 35.91 (SD = 13.59) and CAR+V with 19 (SD = 7.39) is different. This confirms the tendency of the RSME values described above that indicate that video reduces cognitive load. The mean value of CAR+V in the NASA-TLX is lower than the mean value of the paper instruction with 26.25 (SD = 10.95). AAR with a mean value of 34.33 (SD = 14.44) is similar to the mean value of CAR, but AAR+V with 28.91 (SD = 16.32) is higher than CAR+V.

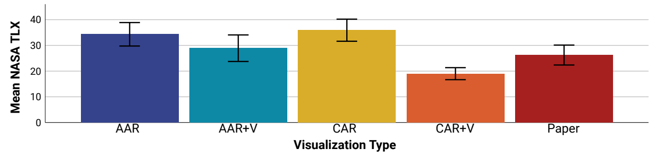

Figure 10: The mean task load during the task using the different visualization types. Error bars represent the standard error of the mean.

6.5 Usability and User Experience

The standard SUS-questionnaire according to Bangor [5] was used to evaluate the usability of the AR visualizations by the participants. The individual scores were combined to determine a mean value and were also tested with a one-way ANOVA. The preconditions of variance homogeneity and normal distribution of the mean values are given. The result of the one-way ANOVA shows no significance between the mean values of the individual visualizations ($F(3, 36) = .483$, $p < .696$). The descriptive statistics provide a mean value of 82 (SD = 12.19) for AAR+V. AAR has a mean value of 79.5 (SD = 9.26), which is only slightly lower. The same is the case for CAR+V with 78.25 (SD = 9.36) and CAR with 76 (SD = 13.5).

The UEQ was used to measure user satisfaction [35]. After completion of the set-up task, the UEQ questionnaire was filled out by the study participants. For the analysis, the standardized evaluation procedure was used and the mean values were determined, as shown in Figure 11. The results show no essential differences between the mean values.

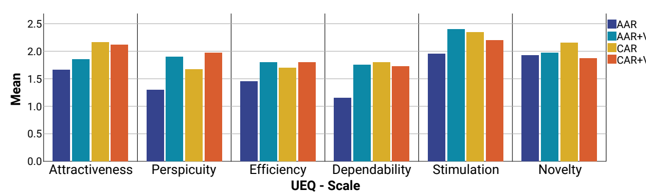

Figure 11: The mean UEQ for the different scales grouped by different visualization types.

6.6 Summary of Qualitative Feedback

Most participants expressed enjoyment in the discussion and closing interview even though half of the participants expressed discomfort related to the weight and comfort of the Microsoft HoloLens. The participants also stated they see high potential for the use of AR for providing assistance in the private and professional sectors. Participant AAR-4 was one of many that gave an example for private use: "*The application was very helpful to me and I can well imagine that it would also support me, for example, in assembling an IKEA cabinet.*" Participant CAR-3 further explained: "*I found it much easier than the paper instructions you always get. Because there I don't understand anything anyway. When I see it virtually, I find it much easier.*" Participant CAR-1 described a potential use case for the professional sector: "*For example, for apprentices or trainees who do not have much experience, they learn how to do it properly. No one would have to stand next to them and explain it, instead they can do it them self, because they have everything in front of their eyes.*"

There were also critical statements. Participant AAR-1, for example, said: "*I noticed that because I was so concentrated on the device and the instructions, I didn't notice important details. You look less at the parts. I think when you work with the AR glasses, you turn off your brain a little bit.*" For similar reasons AAR+V-6 suggested "*I would have found it cool to get feedback from the system. For example, when you placed it correctly, that there is some audio feedback that indicates the part was aligned correctly.*"

The participants using AAR had some issues with identifying the meanings of the abstract 3D models and therefore made the most errors. For this reason, Participant AAR-8 requested additional textual information, AAR-6 asked for concrete presentations of the components, and AAR-5 wanted to see videos about the execution of the tasks. Participants using the concrete AR application also had suggestions with regard to the information visualization. Participant CAR-3 suggested additional simple visualizations like arrows or circles for highlighting important details or positions. CAR-1 asked for the ability to hide the holograms after it became obvious where to mount the component because the holograms blocked his view on the real part. Multiple participants using the videos wished for more detailed videos, for example, with additional highlights of important details with marks or zooms.

Overall, the videos were perceived as extremely helpful by all participants. As AAR+V-1 put it: "*In the video I see how it is done and I just imitate it*". CAR+V-6 explained how she used the holograms and videos: "*I used both equally, because they provided different information. So, the hologram for the position, there I did not look at the part itself and in the video, I was able to get more information about the exact look of the part and where the screws go.*" When we asked the participants, who had used the videos whether they could have completed the task without the video, they expressed skepticism and said that the videos gave them a feeling of safety and strengthened their confidence in their own actions.

7 DISCUSSION

The goal of our study was to examine the influence the visualization of AR content has on complex working practices. We investigated if significant differences exist in (1) task completion time, (2) number of errors and (3) cognitive load between the visualization forms.

7.1 Type of visualization has no impact on completion time

The first hypothesis (H1) is not confirmed. The descriptive statistics show that AAR provides slightly better completion time values than CAR. These results are in line with results presented by Funk et al. [15]. A detailed analysis of our results (see Fig. 6) shows that the standard deviation of the task completion time for paper instructions is much higher than the standard deviation of the task completion time using the other visualization types. Participant CAR-3 suggested that paper instructions are often not very easy to interpret. This statement is reflected in the higher standard deviation. The results that CAR and AAR do not differ significantly in execution time support the findings of Blattgerste et al. [6]. However, the results of Radkowksi et al. [31] differ from our results as Radkowski et al. showed significant differences for task completion times using different visualization types. Also, supplementing AAR and CAR visualizations with videos does not result in significant differences in the execution time. Comparisons of the results between the studies in general are limited, since the studies differ greatly in their design and implementation [37].

7.2 Concrete visualization with videos promotes a correct execution

The second hypothesis (H2) is confirmed as there are significant differences between the individual visualizations with respect to the number of errors. Most participants using CAR performed better than the participants using AAR. Both CAR+V and AAR+V are better than CAR and AAR. In the case of AAR+V, using videos cut the errors in half. Based on these results, it is obvious that the effort to create CAR is worthwhile, as more detailed holograms lead to fewer user mistakes during assembly. AAR should only be used if holograms are combined with a video, however. Only in this case the abstract information of the hologram can be translated into useful instructions for the user [15, 48]. A comparable result is presented by Smith et al. [37], who state that in-situ videos in an AR environment are an effective way to display procedural instructions for assembly tasks. In general, it can be stated that videos complement the holograms and lead to a higher user safety during the execution of the tasks. The participants expressed appreciation of the holograms providing the position and orientation of the tool on the machine. They also stated that if they were unsure about the exact steps required, they watched the relevant video for further information and support.

A more precise differentiation of the results categorized by process step shows that most errors were made in set-up steps 5 and 8 (see Figure 8). The components that had to be mounted in steps 5 and 8 are almost symmetrical and the correct orientation can only be recognized by a small edge on the tool (see Figure 12). In these cases, both visualization types without videos perform poorly, but even though both had cues for the correct orientation. In the case of AAR, an arrow hologram was used to point out specific features (for example the chamfer on component of step 5). The CAR holograms represent the geometric features in the CAD models per se. As the results point out, the cues – especially in the AAR visualization – were not recognized by the participants. Some participants mentioned this aspect during the interviews by asking for a more detailed description of the tool position and orientation in step 5 and 8. We therefore suggest adding textual information for these specific details.

As Radkowski et al. [31] have already shown, set-up steps of varying complexity have a particular effect on the number of errors. The complexity of a single set-up step can be determined, for instance, by the maximum number of orientations of the object to be mounted and the number of contact surfaces which have to be brought into line. Radkowski's et al. [31] results show that CAR visualizations perform better by causing fewer errors. However, the hypothesis that CARs are better suited for complex set-up steps cannot be confirmed. Our results extend these findings by showing that CAR may cause fewer errors the addition of videos is essential when focusing on complex set-up steps. For the complex steps 5 and 8, the instructions with video perform better than the instructions without video. In addition, paper instructions are also quite suitable

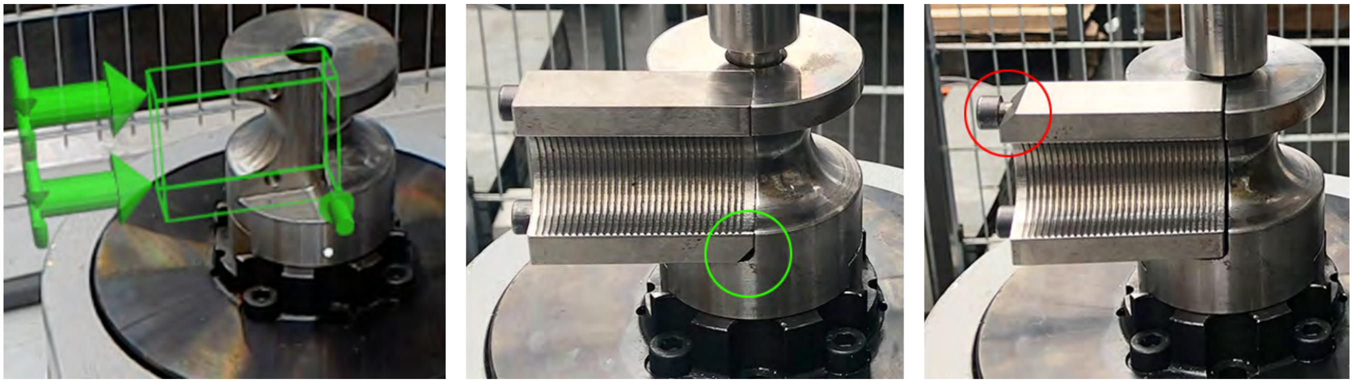

Figure 12: Comparison of the correct mounting of the component in step 5 and a typical error done by the participants. Left: AAR visualization with hint for attention to the chamfer. Middle: correct orientation. Right: the wrong orientation leads to a gap between the components.

for complex steps (see Figure 8). Participant AAR-1 speculates on possible reasons why pure AR instructions do not work so well. In his opinion, the holograms cause the user to be distracted and pay too much attention to the holograms, turning off common sense. Based on this, concrete suggestions were provided during the interviews. For example, participant AAR-8 requested additional written information for particularly critical set-up steps. In addition, the participants would have liked additional feedback on critical set-up steps (AAR+V-6). This suggests that the potential limits for both CAR and AAR were reached, since only the target state is visualized with no verification of the manual assembly process. This could be supported by additional sensory feedback components of the machine. An additional motivation for a subsequent sensory check is the partial lack of accuracy of the holograms. Especially for tasks in which exact placement of the assembly components is important, both CAR and AAR had limitations due to the hardware [6]. Our study does not provide any further insights regarding input interaction mechanisms for HMDs. The literature only refers in one place to the fact that different input interactions have no further impact on performance [37]. However, our qualitative results highlight that participants emphasize further interaction with the AR-device in the form of direct feedback on the performed manual operation in the case for complex assembly operations or to use voice interaction to ease the general interaction with the applications.

7.3 The more concrete the visualization, the lower the cognitive load

The third hypothesis (H3) is not confirmed. The Games-Howell post-hoc test could not identify significant differences between the groups; only the pairwise test of paper and AAR was close to significance ($p < .055$). The task load was the lowest with CAR. The Bonferroni post-hoc test revealed no significant difference between the group, only a trend that CAR+V performs better than CAR ($p < .057$). In their study, Blattgerste et al. also show significantly lower values for paper instructions [7]. Furthermore, AAR+V and CAR+V lead to a decrease in the cognitive load of the participants compared to AAR and CAR. We observed this tendency also in the task load; although they are not statistically significant, they are in line with the findings from the literature [7, 37] and warrant further research. The values for CAR and paper differ greatly between the mental effort and task load and they are not in line with our expectations. Participant CAR-1 reasons the lower task load by the autonomous learning environment that the AR instructions provide. The novices are therefore able to act on their own. In general, this is more successful with CAR compared to AAR and videos in particular reduce the task load considerably.

Both the results of the SUS and the UEQ indicate that there are no problems regarding usability and user experience. The SUS score is above the critical value 68 [5] in all cases considered. The user experience is also rated above average for all visualization forms [35]. The comparison between the visualization forms on the basis of descriptive statistics reveals that AAR is slightly lower than the others. The low values of AAR for "Perspicuity", "Efficiency" and "Dependability" can be traced to the fact that only abstract representations are not sufficient (see AAR-8, AAR-6 and AAR-5).

7.4 Trade-off between accuracy of CAR and applicability of AAR

Currently, paper-based instructions are still status quo in many industrial settings [12]. However, AR-based approaches are increasingly gaining ground in this domain. As participant CAR-3 points out, AR instructions are generally much more accessible than paper instructions. This statement is not consistent with the results from the study and should therefore be judged as a subjective feeling. As our results indicate, the use of paper instructions is certainly justified in simple and complex settings. While in this study, we mainly focused on visualization, the creation of content must also be considered. Content creation requires several skills such as 3D modeling, computer graphics and animations, programming and expertise about the spatial registration of virtual elements and its tracking. Machine set-up experts often have little or no knowledge of creating AR content [17], which means that appropriate AR authoring tools are required to allow a quick and easy way to create content. Methods that involve generating AR content automatically are not suitable in view of the complex tasks [13]. Instructions using AAR can be created directly with the Microsoft HoloLens and none of the listed skills are usually required, as predefined standard sets can be used. However, the results of the study show that the pure abstract representation is not sufficient for real world tasks, and too many errors are made as compared to CAR. Therefore, it must be weighted up whether concrete visualizations are needed, or whether abstract visualizations are sufficient.

7.5 Videos have a high positive effect on abstract visualizations

Although our study has shown that the addition of videos has a positive effect on both visualization types, AAR should always be combined with videos. The combination of AAR and video results in significant improvements and thus meets both requirements; easy and fast creation of AR content, and good performance when using instructions. As participant AAR+V-1 states, the video sequences are well suited to simply being imitated. Moreover, from the participants' (CAR+V-6) point of view, holograms and videos provide different information regarding the set-up task, so that both types of visualization complement each other and do not provide redundancy. Only the increased cognitive load when using AAR+V compared to CAR+V is a disadvantage. With regard to content, Gatullo et al. [17] highlight the updateability of the instructions. Both AAR and video have good updateability if the AR hardware supports the creation of videos.

7.6 Take-aways for designing AR-instructions

Our study shows that when focusing on minimizing the error rate, CAR visualizations are much more appropriate than AAR. When dealing with complex tasks, the error rate is often more crucial than the completion time. We therefore recommend using CAR to avoid errors. However, designers must consider the influence of the AR instructions on the users. Holograms can easily lead users to just focus on the holograms. As adding videos as an additional resource, does not have a negative effect on the execution time, we recommend to integrate additional media like videos and photos or even text if holograms only cannot convey all important information reliable. More strictly, we recommend for both visualization types that videos should always be considered as an essential part of AR-based instructions because videos lower the perceived complexity of a task, they lead to a reduced mental effort during the execution of a complex task and they can reduce errors. We further recommend, whenever, 3D models of machine parts already exist and AR development expertise is available, consider using CAR or CAR+V. Creating CAR and CAR+V require additional effort in preparing the instruction, but for complex tasks this pays off. In the case that no 3D models are available or the creation of the models is too time-consuming, our results show that AAR+V has a performance comparable to CAR+V. The average number of errors is slightly higher, whereas the mean task completion time is lower.

8 LIMITATIONS AND FUTURE WORK

Both the sample size and the set of participants pose a limitation in that the results are not fully representative. A larger sample size as well as a more representative set of participants with different domain knowledge probably would lead to further insights regarding the types of visualizations. To shed more light on the influence of video performance, it would be necessary to adopt a video-only condition in a similar study in the future. Further limitations are the slightly different interfaces of the CAR and AAR applications. The influence of the interfaces and a certain learning effect of the participants during the execution of the study were neglected in the evaluation. Although we assume that many of our findings can be applied to other complex tasks, we cannot say so at this time. Therefore, further research is necessary to investigate other fields of application and, if necessary, to make valid statements about the transferability of the visualization forms. In addition to that we plan to focus on the effort of domain experts in creating the individual types of visualization and whether the effort is justified from their point of view. Although our study provided us with a scenario that was as close to reality as possible, our participants were all inexperienced in handling the machine. In future work, it will be interesting to investigate what effects the respective types of AR visualization have on experts in the field of machine set-up. For example, we assume that experts with significant previous knowledge need fewer concrete visualizations as they already know many aspects of the set-up. Other considerations for further work include investigating the effect of the HoloLens hardware and the users' experience with it. We will further consider multi-modal interactions like voice, hand gestures and gaze and dwell, depending on the task and environment. In addition, the level of detail in the holograms that can be placed on the physical machine to enable a precise representation should be evaluated [6]; this was particularly evident in the execution of step 8 in this study.

9 CONCLUSION

As the use of AR in industrial settings increases, it is important to understand how different visualization types influence user's performance. While previous studies typically either focus on comparing different media [38, 42] as well as AR hardware types [7, 16] or using static apparatus' [31], we evaluate different types of visualization for the same hardware type for complex tasks in mobile settings. To accomplish this, we conducted a comparison study in which we identified the influence of the AR-based visualization types on the performance of complex machine set-up tasks. Both abstract and concrete visualizations have advantages and disadvantages. A designer should always make sure that the choice of elements fits the complexity of the task. Therefore, especially combinations of both visualizations types are useful and there is not only either or. We revealed different lessons learnt for designing AR-instructions using abstract as well as concrete visualizations as well as evaluation criteria.

To conclude our findings, we can state that CAR does not provide a significant advantage to AAR in terms of completion time. In contrast to this, CAR outperforms AAR regarding error rates, and videos lead in both cases to fewer errors – especially for complex set-up steps. Overall, with regard to complex tasks, CAR is the better type of visualization. However, it generates additional work in the course of authoring which should not be ignored in such highly specialized application areas. The need to create time-consuming CAR content can be avoided by complementing abstract visualizations with videos, such that AAR+V is an efficient alternative to CAR or even CAR+V. AR glasses include various on-board features to easily create AAR instructions. If instructions will be created and updated by non AR developers on the shop floor, we recommend using these on-board features to create AAR or AAR+V.

Despite the limitations and the need for further research, our study provides a novel evaluation of different types of visualization on the same hardware and with the same media type for applicability in complex application settings. We not only contribute design spaces for HCI researchers, but we also provide concrete starting points for AR designers and hardware developers.

ACKNOWLEDGMENTS

We would like to thank Arif Atabey, Marcus Schweitzer and Nils Darwin Abele for their assistance, all participants who took part in the study, and the anonymous reviewers for many helpful comments on our work.

REFERENCES

[1] Murat Akçayır and Gökçe Akçayır. 2017. Advantages and challenges associated with augmented reality for education: A systematic review of the literature. *Educational Research Review* 20 (2017), 1–11. https://doi.org/10.1016/j.edurev.2016.11.002

[2] Clemens Arth, Raphael Grasset, Lukas Gruber, Tobias Langlotz, Alessandro Mulloni, and Daniel Wagner. 2015. The History of Mobile Augmented Reality. http://arxiv.org/abs/1505.01319

[3] Doris Aschenbrenner, Meng Li, Radoslaw Dukalski, Jouke Casper Verlinden, Radoslaw Dukalski, Jouke Verlinden, and Stephan Lukosch. 2018. Exploration of different Augmented Reality Visualizations for Enhancing Situation Awareness for Remote Factory Planning Assistance. In *IEEE VR*. 3–7. https://doi.org/10.13140/RG.2.2.14819.66083

[4] Ronald T. Azuma. 1997. A survey of augmented reality. *Presence: Teleoperators and Virtual Environments* 6, 4 (8 1997), 355–385. https://doi.org/10.1.1.30.4999

[5] Aaron Bangor, Philip T. Kortum, and James T. Miller. 2008. An empirical evaluation of the system usability scale. *International Journal of Human-Computer Interaction* 24, 6 (7 2008), 574–594. https://doi.org/10.1080/10447310802205776

[6] Jonas Blattgerste, Patrick Renner, Benjamin Strenge, and Thies Pfeiffer. 2018. In-Situ Instructions Exceed Side-by-Side Instructions in Augmented Reality Assisted Assembly. In *Proceedings of the 11th PErvasive Technologies Related to Assistive Environments Conference*. ACM, New York, NY, USA, 133–140. https://doi.org/10.1145/3197768.3197778

[7] Jonas Blattgerste, Benjamin Strenge, Patrick Renner, Thies Pfeiffer, and Kai Essig. 2017. Comparing Conventional and Augmented Reality Instructions for Manual Assembly Tasks. In *Proceedings of the 10th International Conference on PErvasive Technologies Related to Assistive Environments*, Vol. Part F1285. ACM, New York, NY, USA, 75–82. https://doi.org/10.1145/3056540.3056547

[8] Eleonora Bottani and Giuseppe Vignali. 2019. Augmented reality technology in the manufacturing industry: A review of the last decade. *IISE Transactions* 51, 3 (2019), 284–310. https://doi.org/10.1080/24725854.2018.1493244

[9] Zhaorui Chen, Jinzhu Li, Yifan Hua, Rui Shen, and Anup Basu. 2017. Multimodal interaction in augmented reality. *2017 IEEE International Conference on Systems, Man, and Cybernetics, SMC 2017* 2017-Janua (2017), 206–209. https://doi.org/10.1109/SMC.2017.8122603

[10] Andy Crabtree, Jacki O'Neill, Peter Tolmie, Stefania Castellani, Tommaso Colombino, and Antonietta Grasso. 2006. The practical indispensability of articulation work to immediate and remote help-giving. In *Proceedings of the 2006 20th anniversary conference on Computer supported cooperative work - CSCW '06*. ACM Press, New York, New York, USA, 219. https://doi.org/10.1145/1180875.1180910

[11] Paul Dourish. 2004. What we talk about when we talk about context. *Personal and Ubiquitous Computing* 8, 1 (2 2004), 19–30. https://doi.org/10.1007/s00779-003-0253-8

[12] Timo Engelke, Jens Keil, Pavel Rojtberg, Folker Wientapper, Sabine Webel, and Ulrich Bockholt. 2013. Content first - A concept for industrial augmented reality maintenance applications using mobile devices. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, Portland, OR, USA, 251–252. https://doi.org/10.1109/ISMAR.2013.6671790

[13] Iñigo Fernández del Amo, John Ahmet Erkoyuncu, Rajkumar Roy, Riccardo Palmarini, and Demetrius Onoufriou. 2018. A systematic review of Augmented Reality content-related techniques for knowledge transfer in maintenance applications. *Computers in Industry* 103 (12 2018), 47–71. https://doi.org/10.1016/j.compind.2018.08.007

[14] Michele Fiorentino, Antonio E. Uva, Michele Gattullo, Saverio Debernardis, and Giuseppe Monno. 2014. Augmented reality on large screen for interactive maintenance instructions. *Computers in Industry* 65, 2 (2014), 270–278. https://doi.org/10.1016/j.compind.2013.11.004

[15] Markus Funk, Thomas Kosch, Scott W. Greenwald, and Albrecht Schmidt. 2015. A benchmark for interactive augmented reality instructions for assembly tasks. In *Proceedings of the 14th International Conference on Mobile and Ubiquitous Multimedia*. ACM, New York, NY, USA, 253–257. https://doi.org/10.1145/2836041.2836067

[16] Markus Funk, Thomas Kosch, and Albrecht Schmidt. 2016. Interactive worker assistance. In *Proceedings of the 2016 ACM International Joint Conference on Pervasive and Ubiquitous Computing*. ACM, New York, NY, USA, 934–939. https://doi.org/10.1145/2971648.2971706

[17] Michele Gattullo, Giulia Wally Scurati, Michele Fiorentino, Antonio Emmanuele Uva, Francesco Ferrise, and Monica Bordegoni. 2019. Towards augmented reality manuals for industry 4.0: A methodology. *Robotics and Computer-Integrated Manufacturing* 56, October 2018 (4 2019), 276–286. https://doi.org/10.1016/j.rcim.2018.10.001

[18] Michael Haller. 2004. Photorealism or/and non-photorealism in augmented reality. In *Proceedings of the 2004 ACM SIGGRAPH international conference on Virtual Reality continuum and its applications in industry - VRCAI '04*. ACM Press, New York, New York, USA, 189. https://doi.org/10.1145/1044588.1044627

[19] Sandra G Hart and Lowell E Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload. Advances in Psychology*. Number 52. 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

[20] Gunther Heidemann, Ingo Bax, and Holger Bekel. 2004. Multimodal interaction in an augmented reality scenario. In *Proceedings of the 6th international conference on Multimodal interfaces - ICMI '04*, Vol. 2017-Janua. ACM Press, New York, New York, USA, 53. https://doi.org/10.1145/1027933.1027944

[21] Sven Hoffmann, Aparecido Fabiano Pinatti de Carvalho, Darwin Abele, Marcus Schweitzer, Peter Tolmie, and Volker Wulf. 2019. Cyber-Physical Systems for Knowledge and Expertise Sharing in Manufacturing Contexts: Towards a Model Enabling Design. *Computer Supported Cooperative Work (CSCW)* 28, 3-4 (6 2019), 469–509. https://doi.org/10.1007/s10606-019-09355-y

[22] Lei Hou, Xiangyu Wang, and Martijn Truijens. 2015. Using augmented reality to facilitate piping assembly: An experiment-based evaluation. *Journal of Computing in Civil Engineering* 29, 1 (2015), 1–12. https://doi.org/10.1061/(ASCE)CP.1943-5487.0000344

[23] Jérôme Jetter, Jörgen Eimecke, and Alexandra Rese. 2018. Augmented reality tools for industrial applications: What are potential key performance indicators and who benefits? *Computers in Human Behavior* 87 (10 2018), 18–33. https://doi.org/10.1016/j.chb.2018.04.054

[24] Bettina Laugwitz, Theo Held, and Martin Schrepp. 2008. Construction and Evaluation of a User Experience Questionnaire. 63–76. https://doi.org/10.1007/978-3-540-89350-9{_}6

[25] Riccardo Masoni, Francesco Ferrise, Monica Bordegoni, Michele Gattullo, Antonio E. Uva, Michele Fiorentino, Ernesto Carrabba, and Michele Di Donato. 2017. Supporting Remote Maintenance in Industry 4.0 through Augmented Reality. *Procedia Manufacturing* 11, November 2018 (2017), 1296–1302. https://doi.org/10.1016/j.promfg.2017.07.257

[26] Maura Mengoni, Silvia Ceccacci, Andrea Generosi, and Alma Leopardi. 2018. Spatial Augmented Reality: An application for human work in smart manufacturing environment. *Procedia Manufacturing* 17 (2018), 476–483. https://doi.org/10.1016/j.promfg.2018.10.072

[27] Microsoft. 2020. Dynamics 365 Guides. https://dynamics.microsoft.com/en-us/mixed-reality/guides/

[28] Microsoft. 2020. Dynamics 365 Remote Assist. https://dynamics.microsoft.com/en-us/mixed-reality/remote-assist/

[29] Miwa Nakanishi, Kei Ichiro Taguchi, and Yusaku Okada. 2010. Suggestions on the applicability of visual instructions with see-through head mounted displays depending on the task. *Applied Ergonomics* 42, 1 (2010), 146–155. https://doi.org/10.1016/j.apergo.2010.06.002

[30] Jacki O'Neill, Stefania Castellani, Antonietta Grasso, Frederic Roulland, and Peter Tolmie. 2005. Representations Can be Good Enough. In *ECSCW 2005*, Hans Gellersen, Kjeld Schmidt, Michel Beaudouin-Lafon, and Wendy Mackay (Eds.). Springer Netherlands, Dordrecht, 267–286.

[31] Rafael Radkowski, Jordan Herrema, and James Oliver. 2015. Augmented Reality-Based Manual Assembly Support With Visual Features for Different Degrees of Difficulty. *International Journal of Human-Computer Interaction* 31, 5 (5 2015), 337–349. https://doi.org/10.1080/10447318.2014.994194

[32] Christian Reuter, Thomas Ludwig, and Patrick Mischur. 2019. RescueGlass: Collaborative Applications involving Head-Mounted Displays for Red Cross Rescue Dog Units. *Computer Supported Cooperative Work (CSCW)* 28, 1-2 (4 2019), 209–246. https://doi.org/10.1007/s10606-018-9339-8

[33] Leonardo Rodriguez, Fabian Quint, Dominic Gorecky, David Romero, and Héctor R. Siller. 2015. Developing a Mixed Reality Assistance System Based on Projection Mapping Technology for Manual Operations at Assembly Workstations. *Procedia Computer Science* 75, Vare (2015), 327–333. https://doi.org/10.1016/j.procs.2015.12.254

[34] A. Sanna, F. Manuri, F. Lamberti, G. Paravati, and P. Pezzolla. 2015. Using handheld devices to sup port augmented reality-based maintenance and assembly tasks. In *2015 IEEE International Conference on Consumer Electronics (ICCE)*. IEEE, 178–179. https://doi.org/10.1109/ICCE.2015.7066370

[35] Martin Schrepp, Andreas Hinderks, and Jörg Thomaschewski. 2014. Applying the User Experience Questionnaire (UEQ) in Different Evaluation Scenarios. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Vol. 8517 LNCS. 383–392. https://doi.org/10.1007/978-3-319-07668-3{_}37

[36] S. S. Shapiro and M. B. Wilk. 1965. An Analysis of Variance Test for Normality (Complete Samples). *Biometrika* 52, 3/4 (1965), 591. https://doi.org/10.2307/2333709

[37] Eleanor Smith, Gordon Semple, Dorothy Evans, Kenneth McRae, and Paul Blackwell. 2020. Augmented Instructions: Analysis of Performance and Efficiency of Assembly Tasks. In *Virtual, Augmented and Mixed Reality. Industrial and Everyday Life Applications*, Jessie Y. C. Chen and Gino Fragomeni (Eds.). Lecture Notes in Computer Science, Vol. 12191. Springer International Publishing, Cham, 166–177. https://doi.org/10.1007/978-3-030-49698-2{_}12

[38] Anna Syberfeldt, Oscar Danielsson, Magnus Holm, and Lihui Wang. 2015. Visual Assembling Guidance Using Augmented Reality. *Procedia Manufacturing* 1 (2015), 98–109. https://doi.org/10.1016/j.promfg.2015.09.068

[39] Ajit C. Tamhane. 1977. Multiple comparisons in model i one-way anova with unequal variances. *Communications in Statistics - Theory and Methods* 6, 1 (1 1977), 15–32. https://doi.org/10.1080/03610927708827466

[40] Arthur Tang, Charles Owen, Frank Biocca, and Weimin Mou. 2003. Comparative effectiveness of augmented reality in object assembly. In *Proceedings of the conference on Human factors in computing systems - CHI '03*, Vol. 9. ACM Press, New York, New York, USA, 73. https://doi.org/10.1145/642625.642626

[41] Andrew J. Tomarken and Ronald C. Serlin. 1986. Comparison of anova Alternatives Under Variance Heterogeneity and Specific Noncentrality Structures. *Psychological Bulletin* 99, 1 (1986), 90–99. https://doi.org/10.1037/0033-2909.99.1.90

[42] Antonio E. Uva, Michele Gattullo, Vito M. Manghisi, Daniele Spagnulo, Giuseppe L. Cascella, and Michele Fiorentino. 2018. Evaluating the effectiveness of spatial augmented reality in smart manufacturing: a solution for manual working stations. *International Journal of Advanced Manufacturing Technology* 94, 1-4 (2018), 509–521. https://doi.org/10.1007/s00170-017-0846-4

[43] Andres Vargha and Harold D. Delaney. 1998. The Kruskal-Wallis Test and Stochastic Homogeneity. *Journal of Educational and Behavioral Statistics* 23, 2 (6 1998), 170–192. https://doi.org/10.3102/10769986023002170

[44] T Viljakainen. 2020. *Adoption of Augmented reality solutions in field engineering and maintenance: Drivers and barriers for organizations*. Ph.D. Dissertation. https://www.theseus.fi/handle/10024/336010

[45] X. Wang, S. K. Ong, and A. Y. C. Nee. 2016. A comprehensive survey of augmented reality assembly research. *Advances in Manufacturing* 4, 1 (3 2016), 1–22. https://doi.org/10.1007/s40436-015-0131-4

[46] Stefan Werrlich, Carolin Lorber, Phuc-Anh Nguyen, Carlos Emilio Franco Yanez, and Gunther Notni. 2018. Assembly Training: Comparing the Effects of Head-Mounted Displays and Face-to-Face Training. In *Lecture Notes in Computer Science (including subseries Lecture Notes in Artificial Intelligence and Lecture Notes in Bioinformatics)*. Springer International Publishing, Cham, 462–476. https://doi.org/10.1007/978-3-319-91581-4{_}35

[47] Stefan Wiedenmaier, Olaf Oehme, Ludger Schmidt, and Holger Luczak. 2003. Augmented Reality (AR) for Assembly Processes Design and Experimental Evaluation. *International Journal of Human-Computer Interaction* 16, 3 (12 2003), 497–514. https://doi.org/10.1207/S15327590IJHC1603{_}7

[48] M. L. Yuan, S. K. Ong, and A. Y.C. Nee. 2008. Augmented reality for assembly guidance using a virtual interactive tool. *International Journal of Production Research* 46, 7 (2008), 1745–1767. https://doi.org/10.1080/00207540600972935

[49] F R H Zijlstra and L Van Doorn. 1985. *The construction of a scale to measure perceived effort.* University of Technology.