Eyes on the Task: Gaze Analysis of Situated Visualization for Collaborative Tasks

Nelusa Pathmanathan*
University of Stuttgart

Tobias Rau
University of Stuttgart

Xiliu Yang
University of Stuttgart

Aimée Sousa Calepso
University of Stuttgart

Felix Amtsberg
University of Stuttgart

Achim Menges
University of Stuttgart

Michael Sedlmair
University of Stuttgart

Kuno Kurzhals
University of Stuttgart

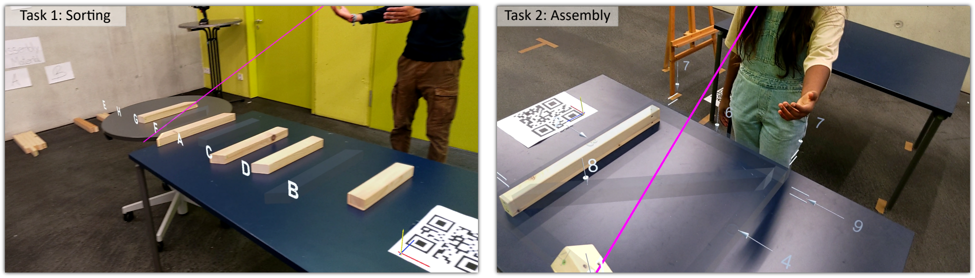


Figure 1: Situated visualizations of timber pieces for (1) a sorting and (2) an assembly task. In both tasks, augmented content supports collaboration partners in performing their tasks without having to consult a paper instruction sheet repeatedly.

Abstract

The use of augmented reality technology to support humans with situated visualization in complex tasks such as navigation or assembly has gained increasing importance in research and industrial applications. One important line of research regards supporting and understanding collaborative tasks. Analyzing collaboration patterns is usually done by conducting observations and interviews. To expand these methods, we argue that eye tracking can be used to extract further insights and quantify behavior. To this end, we contribute a study that uses eye tracking to investigate participant strategies for solving collaborative sorting and assembly tasks. We compare participants' visual attention during situated instructions in AR and traditional paper-based instructions as a baseline. By investigating the performance and gaze behavior of the participants, different strategies for solving the provided tasks are revealed. Our results show that with situated visualization, participants focus more on task-relevant areas and require less discussion between collaboration partners to solve the task at hand.

Index Terms: Human-centered computing—Visualization—Visualization techniques—; Human-centered computing—Visualization—Visualization design and evaluation methods

1 Introduction

Research on augmented reality (AR) interfaces steadily increased and investigated the benefits in various tasks [5, 54]. Examples can be found for navigation tasks [29, 52], sports [34], assembly instructions [11], and data communication between sensor-equipped machinery and people [6]. Collaborative work in AR forms an important subtopic in this research field, for instance, during assembly for fabrication processes [63]. In such cases, human-computer and human-machine interaction contribute to solving the task, and human-human interaction is vital for coordination and efficient performance. Understanding how people perceive and interact with such new interfaces, and also with each other, is crucial for improvement. A critical component of almost all these AR interfaces is situated visualization (SV). SV shows information in the context of the position or state of an object [33] and has been deployed in many scenarios, for instance, for industrial applications [14].

To this point, user-centered experiments with AR interfaces have mainly been evaluated by observations and interviews with the participants [30, 40]. We argue that eye tracking can add a further angle to triangulate our empirical understanding of AR interfaces. Eye-tracking capabilities are already built into head-mounted devices (HMDs) such as the Microsoft HoloLens2 and the Varjo XR, and are thus easily accessible to many researchers. These devices can record the gaze behavior of individual participants for live interaction [37, 49] and post-experimental analysis [12, 47]. Eye tracking-based evaluation has a long tradition in psychology research and is also often applied to evaluate visual stimuli in human-computer interaction (HCI) and visualization research [31], mainly to better understand viewing behavior. In this vein, recording gaze data from participants performing tasks with AR interfaces provides new ways to understand task performance concerning visual attention on virtual and real objects. Hence, the main research question we want to answer is: *How does the use of augmented instructions, such as SV of objects, influence people's behavior in collaborative tasks?*

This question addresses, in particular, explainable measurements for differences in individual performance. If participants perform better with SV support, how is this reflected in their viewing behavior? Vice versa, if participants had issues with SV in AR, can this be explained by investigating their eye tracking data? These questions can partially be addressed by interviews, but eye tracking provides an additional objective measurement to assess this behavior. We focus on non-remote collaborative work in the context of sorting and assembly tasks as they are commonly performed in fabrication and construction processes. This scenario covers a wide range of applications, and we see the proposed approach as a generalizable concept to evaluate user-based experiences with AR interfaces.

Our contributions can be summarized as follows: (1) An eye-tracking experiment investigating the differences in viewing behavior between paper-based instructions (PI) as a baseline, and the presented AR interface that leverages SV. (2) An AR interface to depict and instruct both tasks by SV (Figure 1). (3) Guidelines

*e-mail: nelusa.pathmanathan@visus.uni-stuttgart.com

for the integration of eye tracking in evaluation procedure for AR scenarios. Our results reveal a substantial difference in viewing behavior between the PI and the SV condition. While participants switched their attention in the PI condition between the instructions and the working area, the SV condition allowed them to focus on the working area. Interestingly, this pattern was not observed in all participant pairs and often depended on other factors, such as, previous knowledge, environmental factors, and the team setup. We examined unusual patterns in more detail by including the recorded video data and the data from our questionnaires in our analysis. This analysis confirmed our findings from the gaze data. Furthermore, gaze data provided cues on the working strategies of participants, which we also investigated with additional video analysis.

2 Related Work

For related work, we investigate work on the topics of guidance techniques in AR/VR, AR for collaboration, typical AR evaluation techniques, and evaluation of eye tracking in AR.

2.1 Visual Guidance in AR/VR

AR as well as VR applications, have the advantage of displaying virtual information into space to guide users during specific tasks. This guidance can be simple virtual wayfinding for maps or up to complex instructions for manufacturing tasks. Blattgerste et al. [9] compared AR head-mounted displays, mobile displays, and traditional PI in assembly tasks and noted that AR HMD yields the lowest error rate overall, but PI fairs better when users locate assembly positions. Within AR display methods, Jasche et al. [24] compared abstract and concrete means of visualization within head-mounted AR and found that although concrete AR visualizations do not lead to faster completion time, they do lead to fewer errors. Similarly, Werrlich et al. [65] conducted a study, where they compared paper-based instructions with HMD-based instructions. The results of their study show that HMD-based instructions outperform paper-based instructions. A quite recent work from Tobisková et al. [58] evaluates their AR-based guidance system for manufacturing wooden trusses by using think-aloud protocols and observations. The authors tested their prototype with a variety of stakeholders, which also included professional assemblers. In contrast, we compare PI with SV for sorting and assembly tasks, including common means of questionnaires and video material, but extend analysis with eye tracking to acquire quantitative measurements.

Early research visualized Building Information Models (BIM) onsite for the observation and quality control of ongoing construction processes [51]. Later, the focus shifted towards the support of craftspeople tasks, such as bricklaying of complex walls [23] to speed up construction time and reduce errors in the assembly processes, or learning a craft [18]. Recent studies focus on the collaborative aspect through the coordination of humans and machines, including AR [3, 66]. In addition to HMDs, the use of mobile devices is also often investigated due to their easy accessibility and handling [4]. This line of work inspired the design of our AR interface, and we contribute a study that—for the first time—is primarily based on eye tracking data, which provides additional information in contrast to traditional evaluation methods in AR/VR.

2.2 AR for Human-Human Collaboration

Collaboration becomes important in different work environments where sharing additional information with a group of people can facilitate understanding. Wang et al. [62] present a multi-user collaborative AR system, which allows remote support, collaborative annotation, and editing in industrial processes. Lukosch et al. [36] present some works related to remote collaboration in AR spaces, trying different methods to use AR remotely and guide local workers during physical tasks ([2], [20], [21], [32], [36], [57]). Billinghurst et al. [7] investigate the communication behavior in co-located collaborative AR scenarios. Knoll et al. [27] utilized co-located collaboration in AR for gaming purposes (escape rooms), and Bork et al. [10] took advantage of sharing additional virtual information with multiple people for educational purposes such as anatomy learning. There are different kinds of application areas in which collaborative AR can be applied. Marques et al. [38] introduce a collaborative AR taxonomy, where they categorize such collaborative AR papers based on team, time, task, communication, scene capture and tracking, context sources, level of user actuation, output and augmentation, input modalities, and research.

2.3 Evaluation of AR

Dünser and Billinghurst [17] summarized the problems researchers face when trying to evaluate AR systems. They provide an overview of evaluation methods successfully tested on AR systems and mention a lack of effective methods, which is mainly caused by the variety of AR systems provided and the fact that it is often unclear, who the end-users of the system will be. Based on their review, they identified different categories, which include human perception and low-level cognition, user task performance, system usability, system design evaluation, and collaboration between users. We are focusing on the collaborative evaluation type in our work. For evaluation methods, they differentiate between objective measurements, subjective measurements, qualitative analysis, non-user-based usability evaluation techniques, and informal testing. A more recent work from Merino et al. [40] categorizes the evaluation of AR into seven different categories, from those categories our work focuses on user performance, understanding environment and work practices, team communication, and team collaboration.

Generally, researchers tend to measure objective measurements in conjunction with subjective measurements [53]. However, most AR-based research nowadays uses subjective measurements, as well as qualitative analysis to evaluate an AR system, like questionnaires, video analysis, and interviews. Objective measurements are often restricted to task completion times, accuracy, and the movement path of participants or objects. There are few objective measurements regarding the cognition and perception of the user. To fill this gap, we propose to leverage eye tracking as an additional objective measurement in such studies. As mentioned by Kurzhals et al. [30], the integration of eye tracking data for evaluation can help the analysts assess the visual attention of users and, therefore, make assumptions about their behavior in specific environments.

2.4 Eye Tracking in AR

The analysis of gaze data in 3D environments can reveal how people behave in certain surroundings, and it is possible to analyze their perceptual and cognitive processes through visualizations [28]. Based on these visualizations, analysts can make assumptions and generate hypotheses regarding the design, features, and understanding of a 3D environment. Besides that, analysts can also use these visualizations to evaluate a given design of a 3D environment, for example, in architecture [42], game design [39], and other areas [30]. Blascheck et al. [8] as well as Sundstedt and Garro [56] provide a literature review on visualizations for 3D gaze data. The majority of presented works, however, concentrated only on the visualization of gaze recorded from a physical environment ([47], [43], [45]) or from virtual environments, which are desktop-based ([55]) or in head-mounted virtual reality (VR) displays ([16], [50], [48]). The visualization techniques presented in these different works often consisted of heatmaps to show the distribution of visual attention or scan paths displaying the fixations and saccades of the gaze data or the derived gaze trajectories, i.e., the scanpaths.

For eye tracking in AR, the majority of related work focuses on gaze as an input modality for interaction. Wagner et al. [61] utilize the combination of gaze and hand for selecting items on a 3D UI. Pfeuffer et al. [49] investigated how gaze-adaptive UI spaces can be designed in AR. Lu et al. [35] used eye-tracking data to activate virtual contents within the periphery of the users through the gaze.

The visualization of the gaze point or the gaze ray itself was often utilized during collaborative tasks in AR. Jing et al. [25] investigated different bi-directional collaborative gaze visualizations to show joint attention as well as user intention. There exists only few work for visualizing gaze in AR. Muchen and Tamke [42] investigated the gaze behavior of users in AR for evaluating architectural designs. Their use case should help to asses architectural designs in order to improve them within an iterative design cycle. The work of Öney et al. [67] introduced a visualization tool, which provides the spatial annotation of fixation data recorded in an AR environment. Based on their tool, analysts can define areas of interest (AOIs), label fixations, and examine annotated scan paths. Evaluating AR using gaze data can help to improve UI designs and interactions by supporting existing evaluations. With our work, we extend the focus of this line of research to collaboration behavior derived from the recorded data on how eye tracking can be applied in comparable scenarios (e.g., human-robot collaborations) in the future.

3 Instruction Interfaces

Based on our literature research, we identified a common approach for situated visualization of virtual content in a real-world context. This virtual element supports orienting and positioning a real object according to defined constraints. The traditional and, in this case, the baseline approach is to provide a written and illustrated instruction sheet with step-wise guidance on how to place and assemble individual parts of an object. We designed a case study consisting of two tasks (*Sorting* / *Assembly*) based on our analysis of the state of the art and our experience gained in previous research projects. The *Assembly* task represents a common manual task in the field of timber frame construction. While effort has been made to automate the assembly of beams, studs, and other building components, they are assembled predominately manually by carpenters or joiners by one person or in small teams, depending on the size, weight, and complexity of the building element. Therefore, we selected the assembly of a timber frame as one of our tasks. The *Sorting* task is of high relevance when previously prefabricated unique timber elements are processed automatically by a different machine. After unique beam elements have been machined in a joinery machine, they have to be sorted manually for an industrial robot to pick up [60]. Similar elements like these can be hard to differentiate and require precise measurement. An incorrect order will result in false assembly, potentially causing collision and damage to the workpiece or even the machine. Sorting these timber pieces with visual guidance through AR should help the participants minimize damage.

Baseline: Paper Instructions (PI) For the investigated tasks, we created a one-page instruction sheet consisting of individual steps with textual instructions and annotated illustrations about how to place and assemble the target structure. Figure 2 shows the setup during the experiment. The instructions were designed to contain a textual description identical to the augmented instructions. Such a tangible plan has advantages, as people can point to specific areas and discuss how to proceed.

Situated Visualization Instructions (SV) The main advantage of the augmented approach is a direct spatial reference that helps in understanding where and how to combine parts. We provide a 3D model of the same model depicted in the PI. The model (Figure 2c) is also labeled and the labels are referenced in the instruction text, here, presented as a virtual panel. For single-person use, this panel could be displayed as a head-up display (HUD) element. For the collaborative scenario, we implemented a spatial display that both participants can refer to during their task. With this approach, we enable better communication between them, as pointing gestures become available, similar to the baseline condition.

Implementation We used *Unity* [59] as development platform for this project with support of the *Mixed Reality Toolkit (MRTK)* [41]. Furthermore, 3D models of the sorted and assembled timber parts are imported into Unity and rendered using the Unity standard shaders. MRTK also provides hand tracking for the users to interact with menus, if necessary, and to act as a trigger and an anchor for the rendered schematic instructions on the devices. Our application is separated into three different packages, where two of these packages are used to perform the study and one part is used to analyze the recorded data. For the study, we use a client package on the devices (HoloLenses) that connects to a server package that runs on a standard laptop and is controlled by the conductor of the study. The implementation of the communication patterns is done using the networking library *Riptide* [64]. We further communicate the position of each participant to the other participant´s device through the server to be able to track time spent looking at the partner. The physical AOIs in the environment, are surrounded by 3D cube colliders. Whenever the gaze ray hits the colliders, the corresponding AOI is saved into our data log. For synchronization of the surroundings and AOIs, we use the spatial awareness system of MRTK and combine it with additional spatial anchors represented by QR codes. For the collection of eye-tracking data, we use *ARETT* [26].

4 Eye Tracking Experiment

The main goal of our experiment is to derive common strategies and their related gaze patterns when solving a collaborative task with and without AR support. Hence, we conducted a mixed methods study collecting qualitative and quantitative data from (n=16) participants. They were paired together and we recorded their behavior while solving two tasks with sets of timber pieces. Figure 3 shows the scenario of a pair solving tasks and their corresponding gaze rays. In a post-experiment analysis, we investigated their performance, gaze data, and their subjective user experience. Figure 2a depicts the setup consisting of two tables for the conditions and the accompanying instructions and material. Since the task did not require any specific expertise, we recruited participants from our institute and campus in Stuttgart using flyers and mailing lists. Each session lasted around one hour and the participants were compensated with 12 EUR. The experiment was approved by the ethics committee of the University of Stuttgart and preregistered [46].

4.1 Conditions and Tasks

We conducted the experiment following a within-subject design. The conditions (Section 3) were the formerly described (1) baseline approach with PI and (2) support with SV. To reduce ergonomic risks and coordinate actions as well as increase efficiency and flexibility, we decided to solve the tasks in a team of two. Each group performed two tasks under two conditions. Within each task, the conditions and material for the task were counter-balanced. The tasks were performed in the order depicted in Table 1.

Task 1: Sorting In the material sorting task, we provided the participants with eight pieces of timber elements with a length between 30 cm to 43 cm. Each piece has a unique length and different cutting angles between 0 and 45 degrees at each end. The placement order of all pieces was provided by PI on a scaffold next to the table, or with SV (Figure 2b).

Task 2: Assembling An assembly task involves a combination of three sub-tasks: (1) identifying the timber pieces from a material supply area, (2) placing them in correct locations, and (3) assembling the elements with a cordless screwdriver while adjusting for material tolerances. We provided the assembly order explicitly to enhance the consistency of the data. The assembly task follows that of a conventional and widely used timber frame geometry with butt joints and screw connections. To reduce the difficulty level, pre-drilled holes and threaded inserts were provided in the timber

Table 1: Tasks and trials for the experiment. Both tasks were performed two times, alternating between PI (baseline) and SV with changing instructions and material. The two different combinations in the *Trial* column, indicate in which order the conditions can occur.

Task	**Trial**	**Description**
Sorting	1. PI / SV 2. SV / PI	Participants received eight timber beams with different lengths and cutting angles. They had to order the elements according to the instructions which were presented either on PI, or as SV.
Assembly	1. PI / SV 2. SV / PI	Five timber pieces had to be arranged and assembled with screws at indicated positions. Holes were pre-drilled for this task and one participant was determined at the beginning of the task to use the electric screwdriver, while the other person assisted.

(a) Experiment setup (b) Performing the sorting task with PI (c) First person view of the assembly task

Figure 2: Pictures of the study setup and study tasks. (a) A photo of the experiment setup with two tables, one for PI and one for SV, the material storage in the background, and the scaffold for holding the PI. One additional table on the SV side was needed to fit the whole visualization. (b) Two participants perform the sorting task using the PI, which is clipped to the scaffold. (c) A first-person perspective of the *Assembly* task is shown. On the table, the SV is rendered and the participant looks at the hand pop-up for further instruction details.

elements. Participants were instructed to assign one person to use the screwdriver, while the other assists. The instructions were again presented by a paper sheet (PI) which was also available virtually (SV) in combination with augmented in-place instructions (Figure 2c).

For a more visual grasp of the exact procedure of the tasks, we recommend watching the supplemented video to this work. Additionally, all instructions are provided in the supplemental material.

4.2 Study Procedure

Participants were instructed about the study procedure, data privacy management, and asked to sign a consent form and demographics survey. Each participant was asked to wear a Microsoft HoloLens2 and perform the eye-tracking calibration provided by the system. The participants were also told not to lift the timber pieces too far off the table to ensure a reliable AOI analysis later. During the tasks, communication from the conductors to the participants was kept low. Usually, there were no further instructions needed, but occasionally participants had further questions or needed confirmation of the instructions. In rare cases, participants were informed again not to lift the timber pieces, or to follow the given order of instructions. For *Sorting*, we varied the placement order, so the participants could not learn between conditions. *Assembly* required the participants to build two different but geometrically-similar structures.

4.3 Measurements

To triangulate the qualitative aspects of general behavior patterns, we further measured performance, eye-tracking metrics, and collected subjective feedback from the participants regarding their experience with both conditions.

Performance In terms of task performance, we focused on completion time and correctness. The participants were asked to solve the task efficiently, but not in competition with the other pairs. Furthermore, we investigated performance changes for the first and second trials per task separately to consider potential learning effects.

Eye Tracking To analyze eye-tracking data, we defined AOIs based on the reference objects important for the task. These AOIs comprise (1) the instruction sheet (PI), (2) the hand pop-up (SV), (3) the table they are currently working on (PI+SV), and (4) the head of their collaboration partner (PI+SV). We investigated the scanpaths to identify different viewing strategies when solving the tasks. Further, we analyzed time spans of shared attention during the tasks to gain insight into collaboration strategies.

Subjective Feedback After each trial, the NASA TLX [22] was answered individually by each participant. Further, a feedback questionnaire regarding the comparison and user experience with the conditions was handed out. The full questionnaire and the instructions can be found in the supplemental material.

4.4 Demographics and General Experience

We recorded 12 pairs of participants (24 people) from which 4 pairs were not included in the evaluation due to incomplete data, caused by network connection issues; all results below consider the 8 pairs who completed the study. Of the remaining 16 participants, 10 were male, 5 were female, and 1 preferred not to indicate their gender. Participants were between 18 and 55 years old. Regarding their experience with AR, 11 out of the 16 participants reported having little or no experience. 10 participants reported having some experience with building furniture, while 2 reported having no experience and 2 considered themselves experts in comparable construction tasks.

5 Results

The results are separated into the analysis of performance, gaze behavior, and user feedback, as mentioned before. For statistical analysis, T-tests as well as Wilcoxon tests were executed, accordingly. Details on normality checks can be found in the supplemental

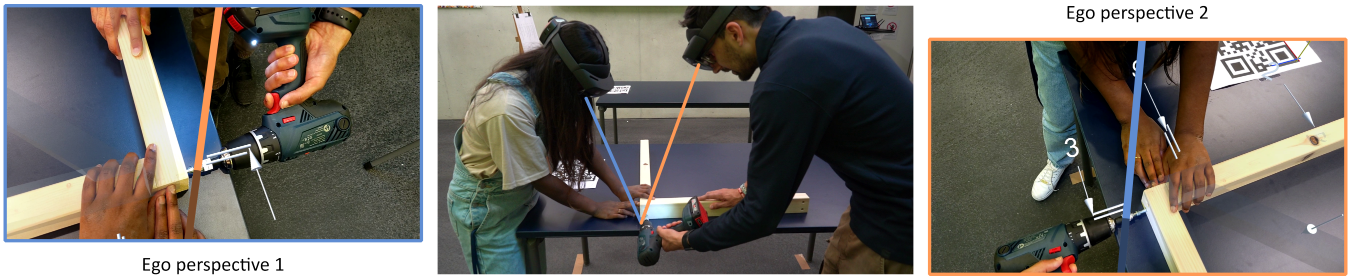


Figure 3: Views of a pair of participants during an assembly task. The picture in the middle shows two participants performing an assembly task, one participant is handling the cordless screwdriver, and the other participant assists by fixing the timber pieces in place. On the sides (left and right) the corresponding augmented first-person view of each participant is shown. Gaze rays are rendered for illustration purposes, they were not visible to the participants.

Table 2: Table of the success rate for *Sorting*. We evaluate the total number of incorrect, the mean (μ) and standard deviation (σ) of correct, and the percentage of correctly placed timber pieces.

Condition	Incorrect	μ	σ	Correct [%]
PI	11	7.31	1.31	91.41
SV	3	7.81	0.39	97.66

materials. Based on an open/axial coding approach with 3 annotators, we extracted strategies from our recorded video content.

5.1 Performance

The participants were able to complete all tasks, although not all sorting attempts were correct.

Correctness The correctness of *Sorting* consists of two measures: order and orientation. These two factors were determined independently, which means that an incorrect order had no influence on the correctness of the orientation. When screening the photos of the results, on the first pass, the order was checked, and in the second pass, the orientation of each individual piece was examined. From these data points, we calculate the sum of incorrect order and orientation, the percentage of correctness, the mean score, and the standard deviation for both conditions (see Table 2). The results show, that the sorting trials with SV were more accurate and less varying than the trials with PI. For *Assembly*, all participants were able to achieve correct result, as it was focused on collaborative behavior, rather than possible errors in construction. Therefore, we did not further analyze the correctness of the *Assembly* task.

Completion Times Figure 4 shows the completion times for all conditions, separated by the trial order. Task completion time improved in all groups between the two trials, indicating a learning effect for the tasks in general. However, groups who started with PI improved more with SV during the second trial — from (μ=330 s, σ=248 s) to (μ=179 s, σ=98 s) in *Sorting*, and (μ=632 s, σ=227 s) to (μ=410 s, σ=75 s) in *Assembly*. We compared the completion times for *Sorting*, when participants started with PI (p = 0.6857) and when participants started with SV (p = 0.1494)(see Figure 4, left side) and for *Assembly*, starting with PI (p = 0.1453) and with SV (p = 0.08647) (see Figure 4, right side). There is a tendency for faster completion times, when participants started with SV instructions. However, we could not identify significant differences regarding completion times between PI and SV.

5.2 Gaze Behavior Analysis vs Open/Axial Coding

Our analysis of gaze data is based on calculated fixations using the I-VT algorithm [44]. For the analysis of the data, we considered the fixation counts, the average fixation duration, and the fixation durations on the different AOIs. Additionally, we investigated the total amount of time participants gazed at the different AOIs in the different conditions. The open/axial coding was conducted using predefined codes based on observed actions. The coders partitioned the videos into 10s segments and assigned codes for each segment, with a total average agreement of 93.21%. We refer to the supplemental material for more details on the coding process.

Figure 5 shows the relative duration of fixations on different AOIs. In *Sorting*, participants focused more on the table with SV (p <0.001), while the focus on the *Instruction* AOI is higher with PI (p <0.001). In terms of gaze duration the participants were looking 23.16% of the time on the table and 6.17% on the instructions for SV and 18.40% on table and 19.90% on instructions for PI.

For *Assembly*, the fixation duration on the table was significantly higher with SV (p = 0.03132), instructions were looked up more often in PI than with SV (p = 0.00346). The total gaze duration here was 64.35% on the table and 5.66% on the instructions for SV and 55.54% on the table and 14.17% on the instructions for PI.

For both tasks, direct gaze at the collaboration partner did not occur often. We see this mainly as a result of the given tasks, which required participants to coordinate objects and their positions rather than having eye contact during conversation.

Figure 6 shows a visualization of the scanpath sequences on the AOIs for the different tasks, commonly denoted as *scarf plots* [8]. The colors in the scarf plots indicate the fixation on predefined AOIs, while the fixation duration is plotted on the x-axis. The scarf plots are grouped by the pairs, each pair contains three rows: The first row indicates durations of shared attention on the same AOI. The second and third rows show the individual fixations of the participants.

The main pattern in all scarf plots is the switch between the instructions (green) and the table (blue) participants are working on. Participants executing PI looked more often at the instruction than those working with SV. We could further identify some characteristic behaviors for both tasks and collaboration strategies.

5.2.1 Sorting

Figure 6 (a) shows how the gaze behavior changes substantially during *Sorting* with PI. The fixations frequently switch between paper and workpieces, where the best performing group (Pair 4) spent little time fixating on pieces, and the slowest (Pair 1) spent a long time looking at one or the other. In the coding, the code *discussion* was most frequently assigned to Pair 1 (μ = 47.67).

Figure 6 (b) shows that during *Sorting* with SV, the pairs (3, 4, 5, 6) who spent less time looking at the hand pop-up instructions were faster in executing the task. The coding further revealed that the pairs (3, 4, 6) had fewer discussions (μ = 1.00) than the slower groups (μ = 10.75). The slower groups spent 20–30% of the time looking at the virtual instructions.

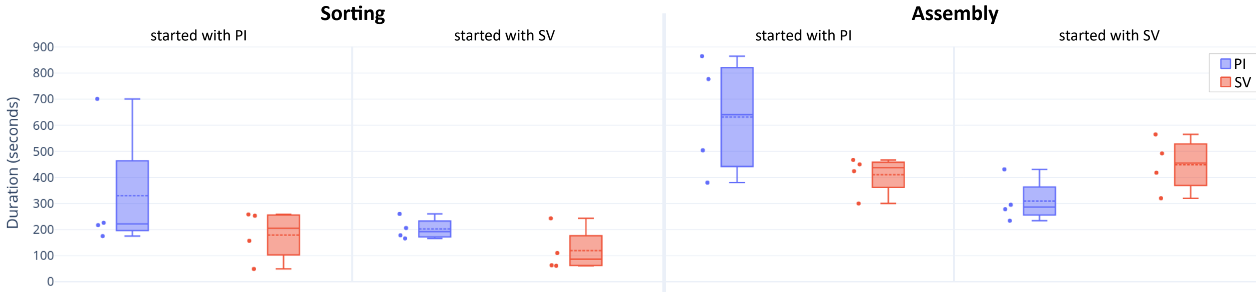
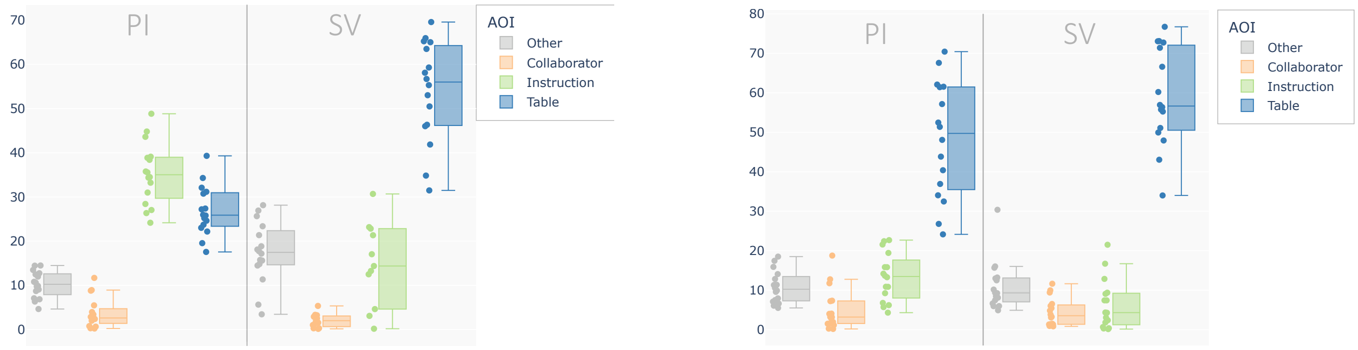


Figure 4: Task completion time between trials represented as box plots with the addition of the individual data points. Each task is split into two trials for PI and SV.



(a) Relative fixation durations on AOIs for *Sorting*

(b) Relative fixation durations on AOIs for *Assembly*

Figure 5: Fixation duration on AOIs during (a) *Sorting* and (b) *Assembly* task with PI and with SV represented as box plots with individual data points next to them. The fixation duration is calculated in percent relative to the task-solving duration. *Sorting* with PI has more fixation durations on the instructions, while SV has more fixation durations on the working area. *Assembly* with PI and SV have the highest percentage of fixation durations on the working area. The fixations durations on the virtual hand pop-up (SV Instruction) is still lower than in the PI task.

5.2.2 Assembly

When fixations on collaborators occurred (orange) during *Assembly*, they mainly resulted from participants in the assisting role (Figure 6 (d), Pair 1). Fixations also reveal the assembly strategy for the timber frame. With PI, Figure 6 (c) shows how pairs (4, 6, 8) first switched between the instructions and the table and subsequently only fixated on the table. During the first phase, they tried sorting the timber pieces according to their order. Therefore, their attention switched between the table and the instructions. After everything was laid out accordingly, they only had to screw the pieces together so fixations were only on the table. In contrast, the other pairs show a shift of attention between the instructions and the table throughout the assembly process. This showed that they were executing the tasks step by step and could further be an indicator of confusion or uncertainty during the task.

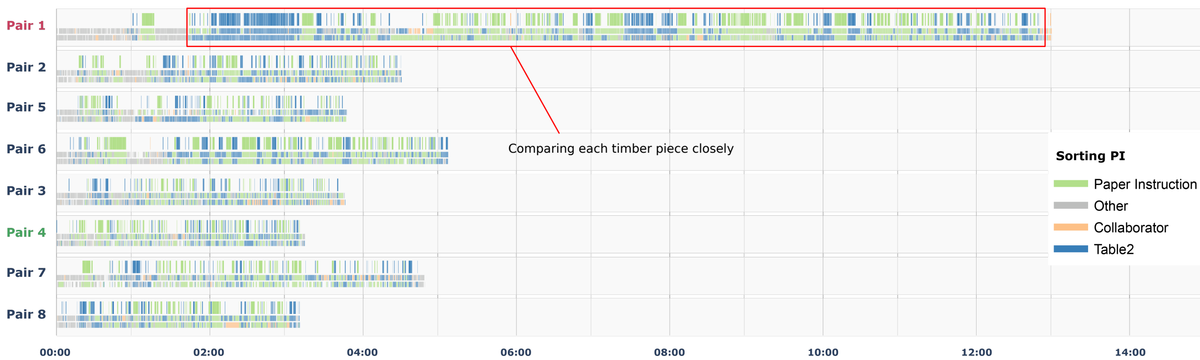
5.2.3 Collaborative Behavior

We further examined collaborative behavior patterns by selecting outlier groups in each condition through the video recordings and cross-referenced this result with the gaze data analysis. The findings are summarized below:

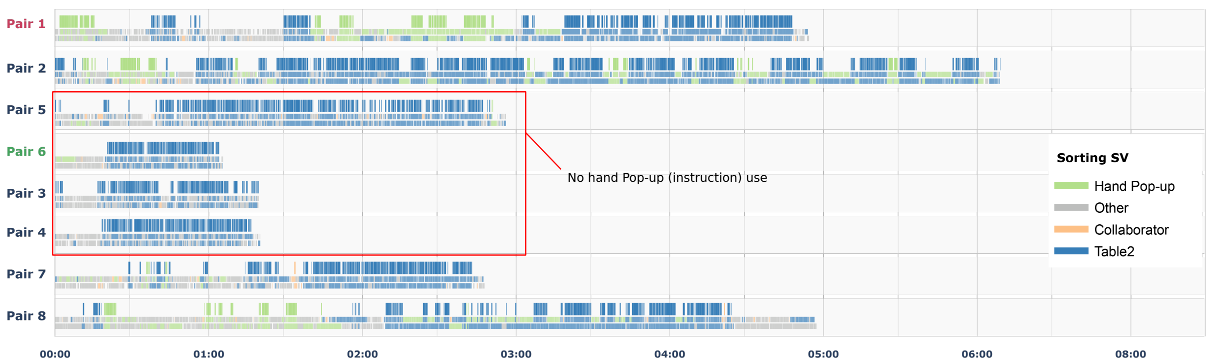
Choice of Strategy The best-performing groups in *Sorting*, regardless of the condition, adopted a relatively ad-hoc strategy. The participants picked up the elements and placed them right away, only checking the result together at the end. In contrast, both groups that over-strategized performed more slowly. For instance, as seen in Figure 6 (a), Pair 1 with PI sorted all the parts by length (along the blue segment) and compared each element closely (alternating green and blue).

Anticipation & Parallel Collaboration During *Assembly*, the screw insertion is a process bottleneck and groups adopted different strategies in this situation. Pair 2 and 7 were, respectively, the slowest in assembly with PI and SV. Neither pair parallelized the screwing process using the screwdriver and the pre-insertion process by hand. Instead of anticipating the next step and helping the partner pre-insert the screws, the participant stood around waiting or did a later step, which does not bypass the screwing procedure as a bottleneck. Because it takes a few seconds to put down the drill, pick up screws, insert screws, and pick the drill back up, this means the average time to insert a set of screws nearly triples (from 10 to 30 seconds) between the fastest and slowest groups. This observation could not be made easily from the gaze patterns, because the participants were either way mostly looking at the workpiece. From the coding results, we noted more instances of collaboration through the code *assistance* in assembly rather than the sorting task (*Assembly* $\mu = 6.79$ and *Sorting* $\mu = 0.35$). Further, the participants moved more often during the assembly than during sorting (*Assembly* $\mu = 3.08$ and *Sorting* $\mu = 0.35$).

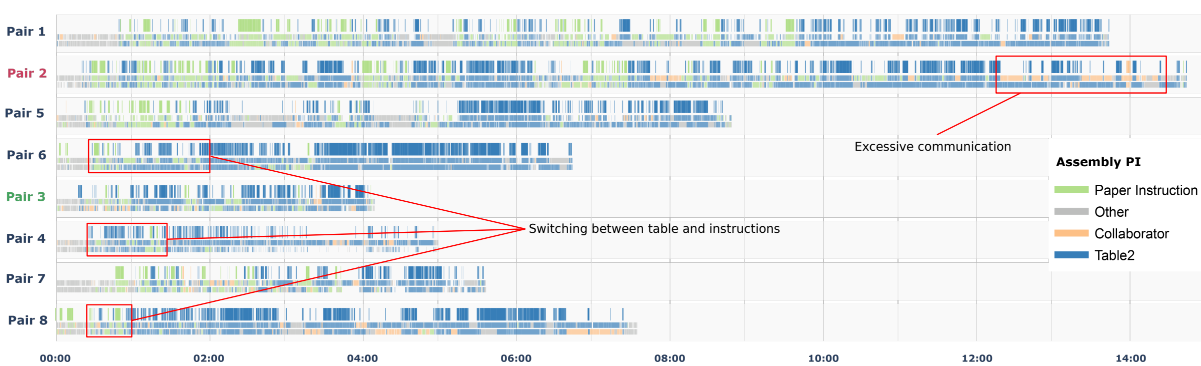
Level of Communication In the slower groups, some engaged in rather excessive communication. They discussed the task and what they could do at length before any action was taken. This can be seen in the gaze pattern where long periods were spent without fixations on the workpiece (e.g., Figure 6 (a) Pair 1 and 6). The coding revealed that during the tasks with PI participants discussed more (*Sorting*: $\mu = 17.58$ and *Assembly*: $\mu = 14.71$). Similarly, some



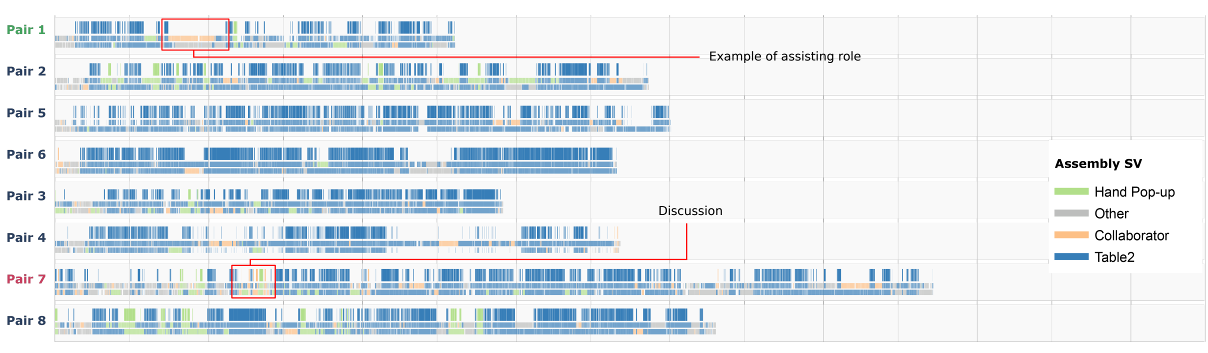
(a) Sorting with PI (Top started with PI)



(b) Sorting with SV (Top started with SV)

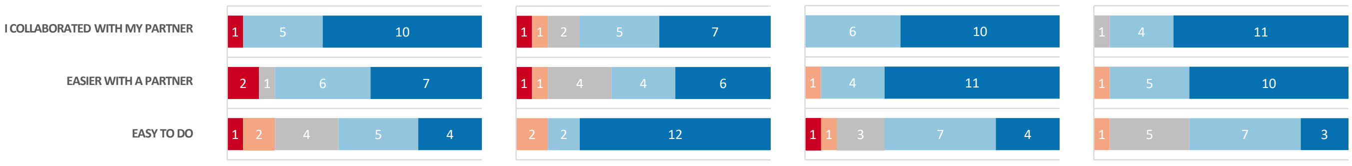


(c) Assembly with PI (Top started with PI)



(d) Assembly with SV (Top started with SV)

Figure 6: Fixation on AOIs during *Sorting* (a,b) and *Assembly* (c,d) tasks with PI and with SV over the task-solving duration in minutes. Each pair block consists of three rows, where the top row shows common AOI intersections and the two smaller bottom rows show the individual gaze data. The top 4 pairs solved the task with one condition first and the other condition second, also vice versa for the bottom 4 pairs. The pairs that have a green text color took the least amount of time to solve the task, while the pairs with red text took the most amount of time.



(a) PI Sorting (b) SV Sorting (c) PI Assembly (d) SV Assembly

Figure 7: Task questions across all conditions and tasks. The charts shows that Sorting tasks felt easier to do than assembly ones. We can also observe that participants collaborated the least when doing *Sorting* in SV.

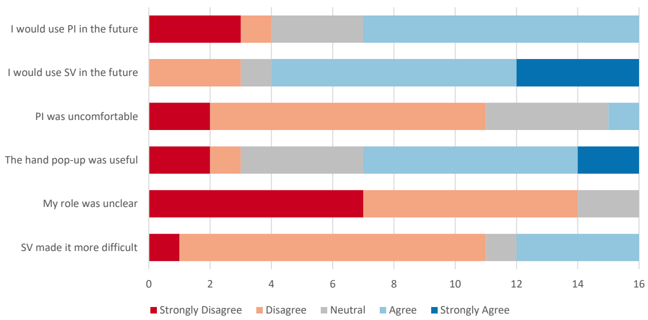


Figure 8: Response of participants on the general feedback questions. We observed that SV did not make the task harder and participants would use it in the future.

pairs did not communicate at all in the first few minutes and each person tried figuring things out by themselves. This is visible from the gaze patterns where there are large gaps in the shared attention (e.g., Figure 6 (b) Pair 2 and 8).

5.3 Feedback

Regarding the task feedback questions, participants rated *Sorting* with SV (12/16) as easier done than PI (9/16). However, they reported slightly more collaboration with their partners when using PI (15/16) than SV (12/16). When looking at these two metrics, we can say that using SV facilitated the users to be more independent when doing the sorting tasks. We also observed that *Assembly* was rated as easier to do with a partner than *Sorting*, which makes sense given the nature of the tasks. Fig. 7 shows the comparisons between the conditions in the two tasks.

We also asked questions about their general user experience in the study after they finished all the tasks (see Fig. 8). Participants pointed out that they would use SV in the future (12/16) more than PI (9/16). This opinion was also reflected in their feedback, mentioning "Situated visualization support the user better to perform the task", "[...] has fast results in construction [...]". Most participants (9/16) agreed that having the hand pop-up was useful. One participant, however, wrote "AR might have been more accessible if not fixed to the hand (thus making its use not hands-free, which can be important for a manual task)". Only a few (4/16) agreed that SV made the task more difficult. Even though the majority would use SV in the future, they mentioned that the design of the SV could be improved through "color-coding", "in-between steps", and "disabling the guidance during mechanical work". Finally, most participants (14/16) reported that their role in the task was clear.

For the NASA TLX data, across all the conditions and tasks, more than half of the participants rated their performance "good". Regarding frustration, at least 10 participants rated it "low"across all conditions, and particularly for *Sorting* with SV 15 participants rated it "low". For Mental Demand, *Sorting* with SV also received lower scores (11/16), while in the other conditions between 6–8 participants rated it "medium low". As expected, *Sorting* received lower scores for Physical Demand than *Assembly*. For *Sorting* tasks, participants also rated the SV condition "low" (12/16) more often than the PI one (5/16). The performance was rated similarly across all the conditions; in the best rating 11 participants graded it as "good" when doing *Sorting* with SV, and in the worst one 9 participants rated it "good" when doing *Assembly* with PI. Detailed charts of the NASA TLX results can be found in the supplemental material.

6 Discussion

To avoid dichotomous interpretations [13, 15] and strengthen replicability, our results comprise traditional statistical methods, extended behavior analysis, and qualitative analysis to address our research question. The presented experiment showed that gaze-based measures provide valuable information for the evaluation of interaction and collaboration in AR. The completion times for *Assembly* showed a non-significant tendency in favor of SV. However, traditional performance analysis (time and error) is not the only relevant aspect in usability and behavior analysis [1, 30], and might even be harmful when considered in isolation [19]. Eye tracking can be used to extend performance analysis. In our case, the eye tracking data showed significant differences regarding fixation durations on AOIs, that it, SV users look less at instructions and focus more on the task, providing new insights into the differences between conditions. If people spend less time on instructions, more complicated tasks might result in better efficiency with SV. Further, the results reveal that the focus of participants is different between *Sorting* and *Assembly*. During *Sorting*, the participants try to sort each piece based on size, shape, and orientation. This caused the participants to check the paper instructions frequently. The gaze data also indicates that, during *Assembly*, the participants seemed to remember the structure after a while and did not have to look at the instructions as often. The qualitative analysis of our data discloses fewer pairs discussing during SV, indicating a higher understanding of the task.

From the scarfplots, we could identify the different working strategies, when it comes to the sorting and assembly of the timber pieces. If the participants followed the instructions step-by-step, this can be identified easily by frequent switching of attention between the respective AOIs. If they tried to figure out the assembly of the pieces first and then screwed everything together, this behavior results in a clearer separation of both AOIs in the scarf plot. Especially for *Sorting* this constant switching between AOIs could be an explanation for the subjectively perceived higher mental demand with PI. If a system's aim is to reduce the mental and physical workload of its users, these results indicate that SV can potentially achieve this by helping users focus on important regions for the task.

The examination of the gaze patterns also allowed the identification of different behaviors between pairs. We could identify which pairs discussed for a long time with their partner, and pairs, which had rather less interaction with their partner. Both aspects are valuable information to include in future techniques to support users with the presented tasks. For instance, an adaptive interface could detect such potential confusion and provide additional cues to solve the task, or vice versa, reduce depicted information in a situated visualization to the essential parts if users know what they are doing.

The gaze patterns show shared attention on AOIs time spans with potentially much collaboration. While specific collaboration behavior still has to be investigated by additional means, for instance, a recorded video of the experiment, this information can substantially reduce search times in the recorded material.

6.1 Open Challenges

We identified some challenges that will require further research. These mainly concern the design of future studies based on our findings and how the evaluation procedure can be extended to dynamic changes during the experiment.

Study Limitations We included inferential statistics to strengthen our assumptions. However, the current experiment is of an exploratory nature as we mainly focus on identifying and describing behavioral patterns during the presented tasks. The descriptive statistics and qualitative observations show tendencies for different strategies that can already help design new interfaces with such strategies in mind. Further, the current experiment investigated the tasks with a learning effect between the first and the second trial. Although the assembly structure was altered, participants tended to be more efficient during the second trial. We noticed that this performance increase was different depending on the starting condition. Combining all these aspects into a single hypothesis-driven study design is nearly impossible and was not the focus of this experiment. Hence, additional experiments with more constrained study designs will be necessary to support the resulting hypotheses. Another limitation of our study is the under-exploration of different visualization techniques for improving situated visualization. The inclusion of highlighting techniques or visualizing instructions stepwise could yield a different performance and different gaze patterns worth investigating and comparing. Besides the visualization, the size of the instruction can also influence the outcome of the study in terms of completion time and correctness. We utilized a DIN A3 sheet for the PI and DIN A4 for the hand pop-up in SV.

Dynamic Visualizations and Environments The current tasks focused on static SV for sorting and assembly. This type of scenario can be generalized to numerous assembly, maintenance, and training tasks. However, tasks with increasing complexity might require animated or interactive situated visualization because one static overview becomes visually cluttered or too complex for efficient interpretation. For instance, we could increase the complexity of the assembly task by adding elements for a 3D structure, which requires additional stabilization and would suffer from typical occlusion problems of 3D visualization. It might also be necessary to adjust visualizations to the changing conditions of the environment or the process itself, e.g., when interacting with robots in collaborative tasks. An evaluation becomes more complicated for such scenarios, especially when environmental changes occur. Dynamic changes in the visualization can be captured programmatically, and gaze hit detections are easy to derive if appropriate collider objects are defined. But changes in the real-world environment have to be tracked separately and might even require manual annotation by investigating recorded gaze replays [30].

6.2 Design Guidelines

Based on our findings, we would like to provide other researchers, who want to include eye tracking in their evaluation of AR scenarios, with some guidance and suggestions on what should be considered when designing SV for collaborative tasks:

Synchronization of Devices: Two AR devices at minimum have to provide synchronized data for analysis. We achieved this by client-server communication, if both devices are recorded individually, additional audio-visual cues should be included for later synchronization. The latter might also be helpful for the inclusion of additional devices, for instance, video cameras for traditional behavior analysis by observation.

Environment Preparation: The presented experiment was conducted in a controlled environment equipped with fiducial markers at important positions for AOI definition. Alternatively, other environment models, e.g., laser scans, could be deployed to match the mesh provided by the AR device with a semantically enriched environment model. Annotation of the mesh from the device would also be possible but might lack details, depending on the applied surface reconstruction techniques.

Statistical and Visual Analysis: We presented an analysis based on common eye tracking metrics such as the relative gaze duration on different AOIs. While this provides an overview, visualization techniques such as the presented scarf plots display a more detailed view of the data over time. If the research question requires a more detailed analysis of specific relations between different AOIs, an analysis by a transition matrix might be helpful.

Following these steps, the integration of eye tracking into the evaluation procedure of AR scenarios is easy to achieve. However, the challenge often remains with the analysis of the recorded data. Hence, we see the development of new techniques and streamlined solutions for analysis in such scenarios as an important and relatively untouched research field for the present and the near future.

7 Conclusion

We investigated how people cooperatively work on sorting and assembly tasks with and without the support of augmented reality guidance by SV. Our results showed that the use of SV tends to improve performance, significantly increases visual attention in the work area, and reduces the need for instructions. These tasks represent a set of basic actions often repeated in numerous fabrication and construction processes. By investigating the viewing behavior, we could identify different strategies that should be considered in future AR interfaces to improve work safety, task efficiency, and reduce mental workload for users to focus on the task at hand.

For future work, we plan to conduct further studies to quantify how common individual strategies occur during the presented and comparable tasks. The results of the included performance analysis are in line with related work comparing situated and paper-based instructions, indicating that SV can help create supportive interfaces and eye tracking analysis showed that a more focused workflow might be one reason for this. While we utilized a scaffold for PI and hand-pop up for SV, different placement of these instructions should be explored in future works. We further want to investigate how the inclusion of live gaze information of the collaboration partner influences the task. In this work, gaze rays were only included for illustration purposes, but might also be a valuable support for coordinating tasks between two or more people working together. Further, it could be worth investigating how gaze patterns differ between AR experts and AR novices.

In conclusion, we want to emphasize the importance of evaluating AR interfaces not just in terms of user experience but also with means such as eye tracking to understand better where issues with an AR interface occur and improve interaction, visual representation, and the experience overall.

Acknowledgments

This work is supported by the Deutsche Forschungsgemeinschaft (DFG, German Research Foundation) under Germany's Excellence Strategy – EXC 2120/1 – 390831618, the Stuttgart Center for Simulation Science (SimTech) under Germany's Excellence Strategy - EXC 2075 - 390740016, and BBSR project 10.08.18.7-22.22. We also thank the International Max Planck School for Intelligent Systems for supporting us.

REFERENCES

[1] Beliv workshop. `https://beliv-workshop.github.io/about.html`, 2022. Accessed: 2024-01-26. 8

[2] M. Adcock and C. Gunn. Using projected light for mobile remote guidance. *Computer Supported Cooperative Work*, 24(6):591–611, 2015. doi: 10.1007/s10606-015-9237-2 2

[3] F. Amtsberg, X. Yang, L. Skoury, H.-J. Wagner, and A. Menges. iHRC: An AR-based interface for intuitive, interactive and coordinated task sharing between humans and robots in building construction. In *Proceedings of the International Symposium on Automation and Robotics in Construction*, ISARC 2021, pp. 25–32, 2021. doi: 10.22260/ISARC2021/0006 2

[4] L. Atanasova, B. Saral, E. Krakovská, J. Schmuck, S. Dietrich, F. Furrer, T. Sandy, P. D'Acunto, and K. Dörfler. Collective AR-Assisted Assembly of Interlocking Structures. In *Design Modelling Symposium Berlin*, Design Modelling Symposium 2023, pp. 175–187, 2023. doi: 10.1007/978-3-031-13249-0_15 2

[5] R. Azuma, Y. Baillot, R. Behringer, S. Feiner, S. Julier, and B. MacIntyre. Recent advances in augmented reality. *IEEE Computer Graphics and Applications*, 21(6):34–47, 2001. doi: 10.1109/38.963459 1

[6] M. Becher, D. Herr, C. Müller, K. Kurzhals, G. Reina, L. Wagner, T. Ertl, and D. Weiskopf. Situated visual analysis and live monitoring for manufacturing. *IEEE Computer Graphics and Applications*, 42(2):33–44, 2022. doi: 10.1109/MCG.2022.3157961 1

[7] M. Billinghurst, D. Belcher, A. Gupta, and K. Kiyokawa. Communication Behaviors in Colocated Collaborative AR Interfaces. *International Journal of Human–Computer Interaction*, 16(3):395–423, 2003. doi: 10.1109/ISMAR.2002.1115083 2

[8] T. Blascheck, K. Kurzhals, M. Raschke, M. Burch, D. Weiskopf, and T. Ertl. Visualization of Eye Tracking Data: A Taxonomy and Survey. *Computer Graphics Forum*, 36(8):260–284, 2017. doi: 10.1111/cgf.13079 2, 5

[9] J. Blattgerste, B. Strenge, P. Renner, T. Pfeiffer, and K. Essig. Comparing Conventional and Augmented Reality Instructions for Manual Assembly Tasks. In *Proceedings of the International Conference on PErvasive Technologies Related to Assistive Environments*, PETRA '17, p. 75–82, 2017. doi: 10.1145/3056540.3056547 2

[10] F. Bork, A. Lehner, D. Kugelmann, U. Eck, J. Waschke, and N. Navab. Vesarlius: An augmented reality system for large-group co-located anatomy learning. In *Proceedigs of the IEEE International Symposium on Mixed and Augmented Reality Adjunct (ISMAR-Adjunct)*, ISMAR 2019 Adjunct, pp. 122–123, 2019. doi: 10.1109/ISMAR-Adjunct.2019.00-66 2

[11] S. Büttner, M. Prilla, and C. Röcker. Augmented Reality Training for Industrial Assembly Work-Are Projection-based AR Assistive Systems an Appropriate Tool for Assembly Training? In *Proceedings of the CHI Conference on Human Factors in Computing Systems*, CHI '20, pp. 1–12, 2020. doi: 10.1145/3313831.3376720 1

[12] S. Castelo, J. Rulff, E. McGowan, B. Steers, G. Wu, S. Chen, I. Roman, R. Lopez, E. Brewer, C. Zhao, et al. ARGUS: Visualization of AI-assisted task guidance in AR. *IEEE Transactions on Visualization and Computer Graphics*, 30(1):1313–1323, 2024. doi: 10.48550/arXiv.2308.06246 1

[13] G. Cumming. The new statistics: Why and how. *Psychological Science*, 25(1):7–29, 2014. doi: 10.1177/0956797613504966 8

[14] L. F. de Souza Cardoso, F. C. M. Q. Mariano, and E. R. Zorzal. A survey of industrial augmented reality. *Computers & Industrial Engineering*, 139:106–159, 2020. doi: 10.1016/j.cie.2019.106159 1

[15] P. Dragicevic. *Fair Statistical Communication in HCI*, pp. 291–330. Springer International Publishing, 2016. doi: 10.1007/978-3-319-26633-6_13 8

[16] A. T. Duchowski, E. Medlin, N. Cournia, A. Gramopadhye, B. Melloy, and S. Nair. 3D eye movement analysis for VR visual inspection training. In *Proceedings of the Symposium on Eye Tracking Research & Applications*, ETRA '02, pp. 103–110, 2002. doi: 10.1145/507072.507094 2

[17] A. Dünser and M. Billinghurst. Evaluating Augmented Reality Systems. In B. Furht, ed., *Handbook of Augmented Reality*, pp. 289–307. Springer New York, 2011. doi: 10.1007/978-1-4614-0064-6_13 2

[18] J. Fong, A. Poon, W. Ngan, C. Ho, G. Goepel, and K. Crolla. Augmenting Craft with Mixed Reality: A Case Study Project of AR-Driven Analog Clay Modeling. In *Proceedings of the 40th Annual Conference of the Association for Computer Aided Design in Architecture*, ACADIA 2020, pp. 436–444, 2020. doi: 10.52842/conf.acadia.2021.232 2

[19] S. Greenberg and B. Buxton. Usability evaluation considered harmful (some of the time). In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, p. 111–120, 2008. doi: 10.1145/1357054.1357074 8

[20] P. Gurevich, J. Lanir, and B. Cohen. Design and Implementation of TeleAdvisor: A Projection-Based Augmented Reality System for Remote Collaboration. *Computer Supported Cooperative Work (CSCW)*, 24(1):527–562, 2015. doi: 10.1007/s10606-015-9232-7 2

[21] P. Gurevich, J. Lanir, B. Cohen, and R. Stone. TeleAdvisor: a versatile augmented reality tool for remote assistance. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, pp. 619–622, 2012. doi: 10.1145/2207676.2207763 2

[22] S. G. Hart and L. E. Staveland. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. *Advances in Psychology*, 52:139–183, 1988. doi: 10.1016/S0166-4115(08)62386-9 4

[23] G. Jahn, C. Newnham, N. van den Berg, M. Iraheta, and J. Wells. Holographic Construction. In *Proceedings of the Design Modelling Symposium Berlin*, Design Modelling Symposium 2020, pp. 314–324, 2020. doi: 10.1007/978-3-030-29829-6_25 2

[24] F. Jasche, S. Hoffmann, T. Ludwig, and V. Wulf. Comparison of Different Types of Augmented Reality Visualizations for Instructions. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '21, pp. 131:1–131:13, 2021. doi: 10.1145/3411764.3445724 2

[25] A. Jing, K. May, G. Lee, and M. Billinghurst. Eye See What You See: Exploring How Bi-Directional Augmented Reality Gaze Visualisation Influences Co-Located Symmetric Collaboration. *Frontiers in Virtual Reality*, 2(1):697367:1–17, 2021. doi: 10.3389/frvir.2021.697367 3

[26] S. Kapp, M. Barz, S. Mukhametov, D. Sonntag, and J. Kuhn. ARETT: Augmented reality eye tracking toolkit for head mounted displays. *Sensors*, 21(6):2234:1–18, 2021. doi: 10.3390/s21062234 3

[27] T. Knoll, A. Liaqat, and A. Monroy-Hernández. ARctic Escape: Promoting Social Connection, Teamwork, and Collaboration Using a Co-Located Augmented Reality Escape Room. In *Extended Abstracts of the SIGCHI Conference on Human Factors in Computing Systems*, CHI EA '23, pp. 1–6, 2023. doi: 10.1145/3544549.3585841 2

[28] M. Koch, D. Weiskopf, and K. Kurzhals. A Spiral into the Mind: Gaze Spiral Visualization for Mobile Eye Tracking. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 5(1):20:1–20:16, 2022. doi: 10.1145/3530795 2

[29] R. Kumaran, Y.-J. Kim, A. E. Milner, T. Bullock, B. Giesbrecht, and T. Höllerer. The Impact of Navigation Aids on Search Performance and Object Recall in Wide-Area Augmented Reality. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '23, pp. 1–17, 2023. doi: 10.1145/3544548.3581413 1

[30] K. Kurzhals, M. Becher, N. Pathmanathan, and G. Reina. Evaluating situated visualization in AR with eye tracking. In *Proceedings of the IEEE Workshop Evaluation and Beyond - Methodological Approaches for Visualization (BELIV)*, BELIV 2022, pp. 77–84, 2022. doi: 10.1109/BELIV57783.2022.00013 1, 2, 8, 9

[31] K. Kurzhals, B. Fisher, M. Burch, and D. Weiskopf. Eye tracking evaluation of visual analytics. *Information Visualization*, 15(4):340–358, 2016. doi: 10.1145/2669557.2669560 1

[32] J.-F. Lapointe, H. Molyneaux, and M. Allili. A Literature Review of AR-Based Remote Guidance Tasks with User Studies. In *Proceedings of the International Conference on Human-Computer Interaction*, HCII 2020, pp. 111–120, 2020. doi: 10.1007/978-3-030-49698-2_8 2

[33] B. Lee, M. Sedlmair, and D. Schmalstieg. Design patterns for situated visualization in augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 30(1):1324–1335, 2024. doi: 10.1109/TVCG.2023.3327398 1

[34] T. Lin, R. Singh, Y. Yang, C. Nobre, J. Beyer, M. A. Smith, and H. Pfister. Towards an Understanding of Situated AR Visualization

for Basketball Free-Throw Training. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '21, pp. 1–13, 2021. doi: 10.1145/3411764.3445649 1

[35] F. Lu, S. Davari, and D. Bowman. Exploration of Techniques for Rapid Activation of Glanceable Information in Head-Worn Augmented Reality. In *Proceedings of the 2021 ACM Symposium on Spatial User Interaction*, SUI '21, pp. 1–11, 2021. doi: 10.1145/3485279.3485286 3

[36] S. Lukosch, M. Billinghurst, L. Alem, and K. Kiyokawa. Collaboration in Augmented Reality. *Computer Supported Cooperative Work (CSCW)*, 24(6):515–525, 2015. doi: 10.1007/s10606-015-9239-0 2

[37] M. N. Lystbæk, P. Rosenberg, K. Pfeuffer, J. E. Grønbæk, and H. Gellersen. Gaze-Hand Alignment: Combining Eye Gaze and Mid-Air Pointing for Interacting with Menus in Augmented Reality. *Proceedings of the ACM on Human-Computer Interaction*, 6(1):1–18, 2022. doi: 10.1145/3530886 1

[38] B. Marques, S. Silva, J. Alves, T. Araújo, P. Dias, and B. S. Santos. A conceptual model and taxonomy for collaborative augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 28(12):5113–5133, 2022. doi: 10.1109/TVCG.2021.3101545 2

[39] N. H. Mat Zain, F. H. Abdul Razak, A. Jaafar, and M. F. Zulkipli. Eye Tracking in Educational Games Environment: Evaluating User Interface Design through Eye Tracking Patterns. In H. B. Zaman, P. Robinson, M. Petrou, P. Olivier, T. K. Shih, S. Velastin, and I. Nyström, eds., *Visual Informatics: Sustaining Research and Innovations*, Lecture Notes in Computer Science, pp. 64–73. Springer Berling, Heidelberg, 2011. doi: 10.1007/978-3-642-25200-6_7 2

[40] L. Merino, M. Schwarzl, M. Kraus, M. Sedlmair, D. Schmalstieg, and D. Weiskopf. Evaluating Mixed and Augmented Reality: A Systematic Literature Review (2009-2019). In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, ISMAR 2020, pp. 438–451, 2020. doi: 10.1109/ISMAR50242.2020.00069 1, 2

[41] Microsoft. Mixed reality toolkit (mrtk), 2022. [Online; accessed 25-May-2023]. 3

[42] Y. Muchen and M. Tamke. Augmented Reality for Experience-centered Spatial Design: A quantitative assessment method for architectural space. In *Towards a new, configurable architecture : Proceedings of the eCAADe Conference*, eCAADe 2021, pp. 173–180, 2021. doi: 10.52842/conf.ecaade.2021.1.173 2, 3

[43] S. Oishi, K. Koide, M. Yokozuka, and A. Banno. 4D Attention: Comprehensive Framework for Spatio-Temporal Gaze Mapping. *IEEE Robotics and Automation Letters*, 6(4):7240–7247, 2021. doi: 10.1109/LRA.2021.3097274 2

[44] A. Olsen. *The Tobii I-VT fixation filter*, 2012. 5

[45] L. Paletta, K. Santner, G. Fritz, H. Mayer, and J. Schrammel. 3d attention: Measurement of visual saliency using eye tracking glasses. In *Proceedings of the SIGCHI Extended Abstracts on Human Factors in Computing Systems*, CHI EA '13, pp. 199–204, 2013. doi: 10.1145/2468356.2468393 2

[46] N. Pathmanathan, T. Rau, X. Yang, A. S. Calepso, F. Amtsberg, A. Menges, M. Sedlmair, and K. Kurzhals. Preregistration link. `https://osf.io/szjpu/?view_only=558ae54e7c4542adb480612ed016aa82`, 2023. Accessed: 2024-01-17. 3

[47] N. Pathmanathan, S. Öney, M. Becher, M. Sedlmair, D. Weiskopf, and K. Kurzhals. Been there, seen that: Visualization of movement and 3d eye tracking data from real-world environments. *Computer Graphics Forum*, 42(3):385–396, 2023. doi: 10.1111/cgf.14838 1, 2

[48] T. Pfeiffer. Measuring and visualizing attention in space with 3d attention volumes. In *Proceedings of the Symposium on Eye Tracking Research and Applications*, ETRA '12, p. 29–36, 2012. doi: 10.1145/2168556.2168560 2

[49] K. Pfeuffer, Y. Abdrabou, A. Esteves, R. Rivu, Y. Abdelrahman, S. Meitner, A. Saadi, and F. Alt. ARtention: A design space for gaze-adaptive user interfaces in augmented reality. *Computers & Graphics*, 95:1–12, 2021. doi: 10.1016/j.cag.2021.01.001 1, 2

[50] Y. Rahman, S. Asish, N. Fisher, E. Bruce, A. Kulshreshth, and C. Borst. Exploring Eye Gaze Visualization Techniques for Identifying Distracted Students in Educational VR. In *Proceedings of the IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, IEEEVR 2020, pp. 868–877, 2020. doi: 10.1109/VR46266.2020.00009 2

[51] J. Ren, Y. Liu, and Z. Ruan. Architecture in an Age of Augmented Reality: Applications and Practices for Mobile Intelligence BIM-based AR in the Entire Lifecycle. In *International Conference on Electronic Information Technology and Intellectualization*, ICEITI 2016, pp. 1–11, 2016. doi: 10.12783/dtcse/iceiti2016/6203 2

[52] S. Rocha and A. Lopes. Navigation Based Application with Augmented Reality and Accessibility. In *Extended Abstracts of the SIGCHI Conference on Human Factors in Computing Systems*, CHI EA '20, pp. 1–9, 2020. doi: 10.1145/3334480.3383004 1

[53] A. Seeliger, L. Cheng, and T. Netland. Augmented reality for industrial quality inspection: An experiment assessing task performance and human factors. *Computers in Industry*, 151:103985, 2023. doi: 10.1016/j.compind.2023.103985 2

[54] S. Shin, A. Batch, P. W. Butcher, P. D. Ritsos, and N. Elmqvist. The reality of the situation: A survey of situated analytics. *IEEE Transactions on Visualization and Computer Graphics*, early access(early access):1–19, 2023. doi: 10.1109/TVCG.2023.3285546 1

[55] S. Stellmach, L. Nacke, and R. Dachselt. Advanced gaze visualizations for three-dimensional virtual environments. In *Proceedings of the Symposium on Eye-Tracking Research & Applications*, ETRA '10, pp. 109–112, 2010. doi: 10.1145/1743666.1743693 2

[56] V. Sundstedt and V. Garro. A Systematic Review of Visualization Techniques and Analysis Tools for Eye-Tracking in 3D Environments. *Frontiers in Neuroergonomics*, 3(1):1–15, 2022. doi: 10.3389/fnrgo.2022.910019 2

[57] M. Tait and M. Billinghurst. The effect of view independence in a collaborative AR system. *Computer Supported Cooperative Work (CSCW)*, 24(6):563–589, 2015. doi: 10.1007/s10606-015-9231-8 2

[58] N. Tobisková, L. Malmsköld, and T. Pederson. Head-Mounted Augmented Reality Support for Assemblers of Wooden Trusses. *Procedia CIRP*, 119:134–139, 2023. doi: 10.1016/j.procir.2023.02.130 2

[59] Unity Technologies. Unity, 2022. [Online; accessed 25-May-2023]. 3

[60] H. J. Wagner, M. Alvarez, O. Kyjanek, Z. Bhiri, M. Buck, and A. Menges. Flexible and transportable robotic timber construction platform – TIM. *Automation in Construction*, 120:103400:1–17, 2020. doi: 10.1016/j.autcon.2020.103400 3

[61] U. Wagner, M. N. Lystbæk, P. Manakhov, J. E. S. Grønbæk, K. Pfeuffer, and H. Gellersen. A Fitts' Law Study of Gaze-Hand Alignment for Selection in 3D User Interfaces. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '23, pp. 1–15, 2023. doi: 10.1145/3544548.3581423 2

[62] J. Wang and Y. Qi. A Multi-User Collaborative AR System for Industrial Applications. *Sensors*, 22(4):13191–15, 2022. doi: 10.3390/s22041319 2

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

[64] T. Weiland. RiptideNetworking, 2022. [Online; accessed 27-June-2023]. 3

[65] S. Werrlich, A. Daniel, A. Ginger, P.-A. Nguyen, and G. Notni. Comparing HMD-Based and Paper-Based Training. In *Proceedings of the IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, ISMAR 2018, pp. 134–142, 2018. doi: 10.1109/ISMAR.2018.00046 2

[66] X. Yang, A. Sousa-Calepso, F. Amtsberg, A. Menges, and M. Sedlmair. Usability Evaluation of an Augmented Reality System for Collaborative Fabrication between Multiple Humans and Industrial Robots. In *Proceedings of the 2023 ACM Symposium on Spatial User Interaction SUI*, SUI '23, pp. 1–10, 2023. doi: 10.1145/3607822.3614528 2

[67] S. Öney, N. Pathmanathan, M. Becher, M. Sedlmair, D. Weiskopf, and K. Kurzhals. Visual Gaze Labeling for Augmented Reality Studies. *Computer Graphics Forum*, 42(3):373–384, 2023. doi: 10.1111/cgf.14837 3