Evaluation of AR Pattern Guidance Methods for a Surface Cleaning Task

Peer-reviewed author version

CEYSSENS, Jeroen; JANS, Mathias; ROVELO RUIZ, Gustavo; LUYTEN, Kris & DI FIORE, Fabian (2024) Evaluation of AR Pattern Guidance Methods for a Surface Cleaning Task. In: Spencer, SN (Ed.). 30th ACM Symposium Virtual Reality Software and Technology, VRT 2024, ASSOC COMPUTING MACHINERY, p. 1 -11.

DOI: 10.1145/3641825.3687730
Handle: http://hdl.handle.net/1942/44844

Evaluation of AR Pattern Guidance Methods for a Surface Cleaning Task

Jeroen Ceyssens
jeroen.ceyssens@uhasselt.be
UHasselt - Flanders Make
Expertise Centre for Digital Media
Diepenbeek, Belgium

Mathias Jans
mathias.jans@uhasselt.be
UHasselt - Flanders Make
Expertise Centre for Digital Media
Diepenbeek, Belgium

Gustavo Rovelo Ruiz
gustavo.roveloruiz@uhasselt.be
UHasselt - Flanders Make
Expertise Centre for Digital Media
Diepenbeek, Belgium

Kris Luyten
kris.luyten@uhasselt.be
UHasselt - Flanders Make
Expertise Centre for Digital Media
Diepenbeek, Belgium

Fabian Di Fiore
fabian.difiore@uhasselt.be
UHasselt - Flanders Make
Expertise Centre for Digital Media
Diepenbeek, Belgium

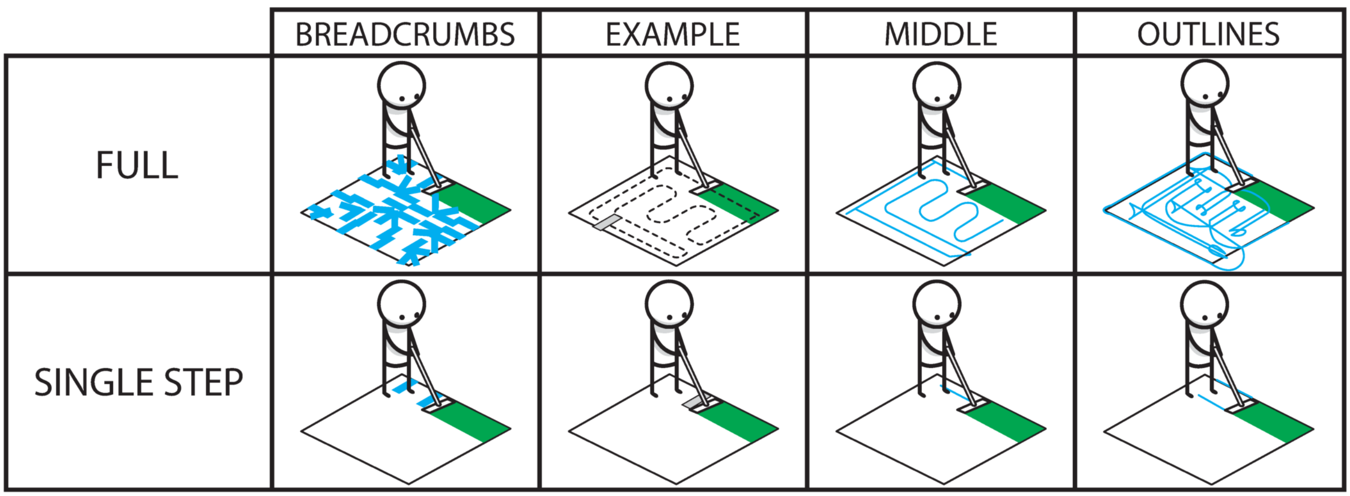


Figure 1: In our evaluation of an augmented reality cleaning application, we tested four types of visual instructions: breadcrumbs, ghost examples, paths, and outlines. These instructions were provided in either full or as a single step.

ABSTRACT

Cleanroom cleaning is a surface coverage task where the pattern should be followed correctly, and the entire surface should be covered. We investigate the efficacy of augmented reality (AR) by implementing various pattern guidance designs to enhance a cleanroom cleaning task. We developed an AR guidance system for cleaning procedures and evaluated four distinct pattern guidance methods: (1) breadcrumbs, (2) examples, (3) middle lines, and (4) outlines. We vary the instructions on the entire surface or as a single step. To measure performance, accuracy, and user satisfaction associated with each guidance method, we conducted a large-scale (n=864) between-subjects study. Our findings indicate that single step instructions proved to be more intuitive and efficient than full instructions, especially for the breadcrumbs. We also discussed the implications of our results for the development of AR applications for surface coverage and pattern optimization.

CCS CONCEPTS

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

KEYWORDS

Augmented Reality; Pattern guidance; Motion control.

1 INTRODUCTION

Virtual reality (VR) and augmented reality (AR) have gained traction across various sectors for training [5, 30, 35] and assistance [9, 13]. These technologies are increasingly being used to perform a wide range of manual tasks and to train dexterity skills. However, guiding continuous movement and supporting fine-grained dexterity skills has proven to be cumbersome and has gained interest from the research community. To address the concern of conveying clear and effective instructions to the user, previous work explored the most efficient type of instruction to guide users in motion activities in terms of efficient execution [4, 15, 18, 31]. We build upon these works and further investigate how to guide manual labour activities that cover an entire surface, such as room painting, spray painting, vacuuming, cleaning, sanding, and plastering, among others.

This paper focuses on motion adherence by studying the efficacy of different instruction types (listed on Figure 1) to allow for full surface coverage. We specifically explore the guidance for a proof-of-concept AR cleanroom cleaning use case. It is important to cover the entire surface area within cleanrooms to avoid contamination of the processes performed in the room [24]. To be able to cover the surface completely, the pattern that users need to follow with the cleaning mop will have to go across the entire surface and require user movement during the activity (see Figure 1). These patterns are hard to continue abiding by for new cleaners, so we believe AR instructions can support them during this operation.

We present the following contributions:

- Design, implementation and validation of a proof-of-concept AR-based surface coverage guidance and progression system.
- A large-scale study to understand the efficacy of commonly presented motion guidance instructions in achieving full surface coverage.
- A set of validated visualization patterns on how to present instructions for manual labor activities, either fully or only when necessary, and how this affects performance and usability.

We first analyze existing research on motion guidance and skill training, and focus on use cases that use AR for floor treatment tasks. We then present the design of our system, which is done in collaboration with cleanroom cleaning experts, and detail the coverage implementation and the four pattern guidance methods we considered (presented in Figure 1). Finally, we present the design of our study, the measured results, and discuss their implications.

2 RELATED WORKS

Our work explores using AR to guide users in surface coverage tasks, aiming to cover an entire surface with an optimal pattern in a single motion, avoiding repetition. Previous studies have examined extended reality (XR) enhancements for guiding users during motion and training new skills. We review these studies and their relevance to our investigation of pattern guidance on surfaces.

2.1 XR Motion Guidance

For sequential path-following tasks (one motion after the other), Liu et al. [15] studied how precues of the next motion in the sequence and glyphs can be used to increase the intelligibility of the path to follow. They found that adding solid path visualizations was most efficient for guidance; however, precues efficiency dropped after four sequence levels. If no path guidance was present, this performance drop would already occur for levels higher than one, indicating their need. In a later work, Liu et al. [17] also explored how goal-based and action-based instructions, together with precues, help users achieve required translations and rotations in sequential motion tasks. Goal-based instructions were found to be most efficient to support users in achieving tasks correctly. However, precues were not useful for most of the participants, highlighting the need to study the effects of precueing upcoming steps more in-depth. Within our work, we specifically study instructions within continuous motion, where rather than smaller sequential tasks, we have one long pattern of continuous motion that needs to be followed as accurately as possible. However, we extend upon the work of precues by providing more or less information on the full pattern depending on the current state of the task.

We are not the first to explore pattern guidance for tasks executed on top of a surface area. For motion skills such as calligraphy, Yang et al. [30] proposed a system called *Just Follow Me* where a ghost representation in a HMD shows how specific motions should be carried out so that users can simply follow them. Similarly, Nomoto et al. [20] studied how adding visuo-haptic feedback can help support a precise motor control task, such as drawing, by moving a dummy hand toward the correct target to force users to adjust their movements accordingly. Prior work of Narita and Matsumaru [19] has looked at enhancing calligraphy-stroke learning using augmented reality projection, where a teacher's brush stroke is mirrored to the participant to replicate the calligraphy pattern. In general, it was found that adding the teacher's brush strokes improved the participant's calligraphy results. In terms of task speed, Ceyssens et al. [4] studied how different AR instructions would help users achieve the correct speed of motion for a line-tracing task. They presented static and dynamic instructions in the form of glyphs, graphs, and an example orb similar to the teacher's brush stroke of Narita and Matsumaru [19]. They found that only the example orb would cause participants to achieve the correct target speed. When looking at live guidance of full-body motions, Yu et al. [34] explored in three studies how adding path guidance helps communicate the desired postures and movements to users. They found that first-person instructions outperformed those shown outside of the participant's body (such as a mirror, third-person, and top-down view), where simple path glyphs were presented for the arm movement. They also found that providing a ghost arm to replicate the desired movement of the body outperformed providing instructions, both in terms of conducting the correct translation, motion pattern, and speed of movement. Within our work, we do not focus on communicating speed to the user; we primarily focus on pattern guidance on how it improves the accuracy of surface coverage.

2.2 Skill Training with XR

XR has been widely explored to train users to acquire new skills or knowledge [11]. In operation contexts, the entire spectrum of mixed reality has extensively been used to train operators to perform new procedures in manual assembly operations [5], with many use cases of virtual reality [29] and augmented reality [1, 27] previously explored. When it comes to training specific skills with serious games, Backlund et al. [2] proposed a system called "Sidh" to train firefighters for fire extinguishing scenarios by projecting fire onto a digital CAVE around the user. Lerner et al. [14] looked at how VR can be used to simulate emergency medical care scenarios and train people in how to perform certain medical procedures. For motor skill training, Ricca et al. [23] studied whether adding the digital hand within VR improved the task performance and usability of a pick-and-place type task, compared to only visualizing the tool interactions. They found no performance difference between the digital hands and no hands conditions, however participants preferred seeing the hands to get a better overview of the procedure. For gesture training, Jeanne et al. [10] studied how 3D visual cues can be used to guide users along the correct path of a gesture and teach them how to perform it correctly. They found that providing a path during the training phase to follow provided better results during the activity itself but worse than their proposed guidance *EBAGG* in the post-training phase where no more guidance was given. Looking more in-depth at the visual feedback techniques for bare-hands interaction in VR, Vosinakis et al. [26] compared different feedback visualizations within VR to traditional desktop methods for grasp-and-release tasks. They found that coloring objects in VR gave more clarity on the interaction than using connecting lines, halos, or shadows, with user performance also being higher in VR.

2.3 Augmented Reality Cleaning

There have been some examples of augmenting cleaning operations within XR. Recently on the consumer market, Dyson unveiled plans for the "CleanTrace" app to highlight in AR where users have vacumed[1]. Within research, Yu et al. [32] used a Microsoft Hololens 2 to visualize contaminated spots and highlight (using hand-tracking) the areas the user has already cleaned using a cloth in hand. Another example is by Fukawasa and Nakayama [7], who used projection augmented reality onto a floor to clean, to build a guidance system for cleaning that specific surface area with a cleaning mop, similar to what we propose. They also presented several instruction areas on top of the surface to indicate the pattern that should be conducted for each area (such as wiping, scrubbing, and sweeping). Compared to their works, we are particularly interested in the exploration of surface cleaning guidance for professional contexts, where a singular motion is ideal to perform the task (compared to repetitive motions).

3 AUGMENTATION OF CLEANING OPERATION

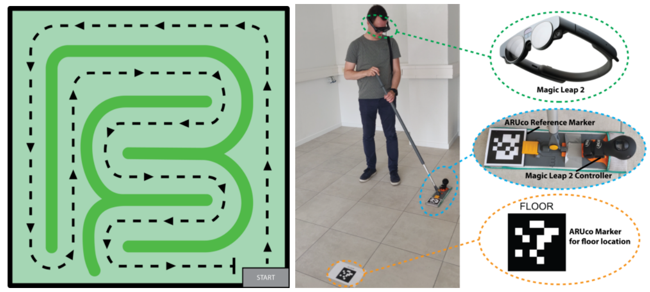
Here we describe the design and implementation of our proof-of-concept enhanced cleaning operation with AR. We also discuss the principles used in the design of our pattern guidance methods to guide users toward the correct adoption of the practices.

3.1 Understanding the problem

To get a better understanding of the needs within a cleaning operation, we interviewed two cleanroom cleaning companies that train and guide cleaning operators. These companies gave us an overview of the standard cleaning procedures within cleanrooms and what important steps they teach to new operators. From these interviews, We identified the following measures of importance to consider within our guidance system:

- Cleaning should always occur according to previously established patterns (see Figure 2a) according to room size to ensure you end back where you started.

[1]https://www.dyson.com/discover/innovation/new-machines/dyson-cleantrace (Last Accessed: 12/08/2024)



(a) (b)

Figure 2: The pattern and implementation with (a) the cleaning motion (black arrows) used to ensure proper overlap, the green area shows what is cleaned, and dark green lines represent overlap, (b) shows the full setup of the implementation.

- During cleaning according to the pattern, overlap should occur between the previously cleaned area and the newly cleaned area between 10 and 20% (see darker lines between the arrows on Figure 2a).
- Cleaning should always happen backward to ensure no walking over cleaned surfaces.
- The mop should always move in one direction to ensure dirt is not left on the sides or behind the mop from build-up.
- Cleaning should not happen too fast to allow for proper soaking of the solution, and not too slow to take forever to clean one room (ideal speed around 0.3 meters per second).

The cleanroom cleaning companies also highlighted the difficulty for (new) cleaners to be able to follow these measures without guidance, especially covering the entire surface accurately in a singular motion and knowing when and where to turn exactly. This motivates a huge part of our research on providing assistance for surface coverage tasks.

3.2 Apparatus & Tracking Methodology

A full overview of the hardware and setup we used to create our guidance system and study can be seen on Figure 2b. To develop an AR enhancement of the cleaning operation, we made use of the Magic Leap 2[2] AR glasses. These glasses use inside-out tracking with an internal ToF depth sensor and RGB-D camera to perform continuous SLAM (Simultaneous Localization and Mapping) of the environment around the user. To track the cleaning mop during the cleaning procedure, we 3D printed a mount for the Magic Leap 2 controller, which uses inside-out tracking with IR LEDs and an onboard lightweight SLAM tracker. One downside of the controller's inside-out tracking is potential deviations in the positional tracking when the primary HMD device does not recognize the controller and sync the SLAM tracking. To balance these issues, we attached an ARUco marker [8] on the other side of the cleaning mop, which is used as a reference marker for the positional tracking (optimized settings for slow and accurate tracking). To build the application, we made use of Unity, which is directly supported by Magic Leap 2.

[2]Magic Leap 2: https://www.magicleap.com/magic-leap-2 (Last Accessed: 12/08/2024)

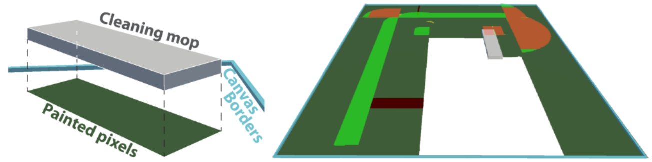
3.3 Surface Coverage Highlights

3.3.1 Design of Surface Coverage. To support operators during the cleaning operation, we created an initial design of the live feedback mechanism together in discussion with the companies. To highlight the area to clean, the idea is to showcase a heatmap of the surface area to clean, where green highlights whether you have cleaned correctly, and red indicates that a mistake has occurred there (e.g., cleaned too fast, cleaned with the wrong side of the mop, walked over the cleaned area). Whenever mistakes occur, a warning should be given to users in their field of vision so they can correct these errors faster. A second heatmap is shown on top of the previous surface to highlight which areas have been cleaned multiple times (such as the required overlap). Using this augmentation, it becomes possible to get an idea of the areas that have been cleaned, what areas have been missed, where overlap happened, and where mistakes have occurred.

3.3.2 Implementation of Surface Coverage. For the heatmap implementation, we made use of a quad in Unity, which serves as the "painting canvas" for cleaning the floor area. An example of this heatmap quad in Unity can be seen on Figure 3b, where the mop is shown cleaning the areas and showing the overlaps as discussed in subsubsection 3.3.1. The canvas is 2 x 2 meters in real life. To paint the pixels of the canvas, we cast the corners of the cleaning mop onto the quad (see Figure 3a) and fill in all the pixels between the corners and the contour of the mop on top of the canvas. We also keep track of every pixel of the canvas, whether it is considered "cleaned", "uncovered", or "faulty" (when going too fast or cleaning in the wrong direction). The color of each pixel is then decided based on their value: green for "cleaned", transparent for "uncovered", and red for "faulty". Cleaning a "faulty" pixel with the correct cleaning method transitions it to a "cleaned" pixel. A second heatmap floats slightly above the original canvas to show the overlap of cleaned areas. All pixels that are cleaned a second time are communicated over to this second heatmap and are then considered as overlapped pixels. An algorithm runs continuously across the second heatmap to identify clusters of covered pixels, decide the width and height of these clusters (based on the shape), and decide the level of overlap and the color to give it. A yellow color indicates an overlap of less than 10%, light green indicates an overlap between 10 and 20%, and orange indicates an overlap above 20%. Finally, the boundaries are visualized in cyan to inform users what area can be cleaned. To optimize the processing speed without sacrificing too much information, the canvas has a texture of 512x512 pixels.

3.4 Pattern Guidance Methodology

Only providing highlights of covered surface area would not be sufficient in professional environments, where failing to adhere to patterns correctly can result in poor task execution in cleanroom environments. We need additional conditions to be met, such as covering a large surface area in a single motion, ensuring users cover the entire surface, and introducing proper overlap with previously cleaned areas. However, we currently do not have an understanding of which visualization type is best to guide users to meet these conditions. To address these concerns, we study several designs



(a) Cleaning mop **(b) Cleaning the platform with the mop**

Figure 3: Surface coverage implementation, with (a) the cleaning mop projection onto the platform and (b) the mop cleaning the platform where green shows what has been cleaned correctly, red where cleaned wrongly (in this case too fast) and yellow, light green or orange show overlap on top.

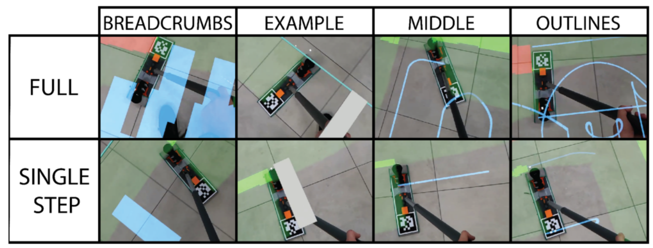


Figure 4: The visualizations tested based on the methodology of breadcrumbs, ghost example, middle line, and outline. "FULL" visualizations cover the entire surface, and "SINGLE STEP" visualizations only show the next part of the pattern.

based on literature and game design practices to visualize a pattern and guide users in executing the full pattern (see Figure 4).

One common design seen in guidance visualization consists of "single step" instructions, a subset of the full instructions that only shows the information required to perform the current step, and showing little to no information on the future steps that are to be performed [12, 21, 28]. Single step instructions lower the distractions of the user by lowering the amount of information (clutter) during the task [16]. This makes single step instructions a viable option to perform motion guidance for surface coverage, where focus on the task at hand should be optimized. To test the efficacy of our guidance visualizations, we consider both the full overview and single step versions, which are described below.

3.4.1 Breadcrumbs. To guide users along a path, in game design and literature, trails are often created with repeated visuals for users to follow toward a specific target area. This concept is called "breadcrumbs" or "footprints"[25]. In our case, we consider breadcrumbs to highlight the pattern to follow by showing intermediate points that need to be achieved (see BREADCRUMBS on Figure 4). When users follow the pattern correctly, the breadcrumbs disappear as they are passed over with the mop. In our design, every breadcrumb shows the ideal orientation and position of the cleaning mop at that point in time and is the same size as the original mop. For every area in between, the user must try to stay at the correct location with no guidance until the next breadcrumb. We expect this should allow them to focus on the cleaning activity since no instruction is visible between breadcrumbs. The breadcrumbs cover the entire canvas for the full version of this instruction. For the single step version, only the next step is shown. When the mop approaches and becomes close to achieving the step, the step after the next will also be shown preemptively.

3.4.2 Example. For this instruction, we make use of the "ghost" metaphor [4, 19, 30], which is an example mop that carries out the pattern correctly for the user to see. In the full version of this instruction, the example continuously moves across the surface area, and users have to follow the movement of this example. For the animation speed of the example, we chose the ideal cleaning speed 0.3 m/s mentioned by the cleaning companies we contacted, which adds an additional layer of information (correct speed). For the single step version of the instruction, the example will always be right in front of the cleaning mop. When the cleaning mop moves correctly, the example will also move forward. If the cleaning mop stops moving, the example also stops. The example instructions are shown under the category EXAMPLE on Figure 4. Since the example is there as a reference to use during the cleaning procedure, we expect that users can copy its movement of whenever they see it to get the ideal result.

3.4.3 Middle. To provide guidance on paths to follow, the most intuitive, well-known, and studied instruction is a line in the middle that directly shows the motion to follow (the MIDDLE instruction of Figure 1). In most of the literature, this instruction is primarily used for tasks with small motions, often consisting of a singular straight line [15, 18, 34]. Our guess is that using only a singular line to inform the pattern to follow makes users potentially more lenient on how accurately they keep the line in the middle of the mop and forget corners or edges as a result. For the full version of the instruction, the line shows the pattern across the entire surface. The single step version shows the middle line for the next 0.6 meters (2 seconds according to the ideal speed).

3.4.4 Outlines. For our final instruction method, we test an alternative to the MIDDLE instruction, where instead of in the middle, we show two lines to represent the outer edges of the cleaning mop (see OUTLINES on Figure 4). With this instruction method, users can balance the cleaning mop between the edges and keep a simple overview of how much the mop deviates from the intended path. The full version of this instruction covers the entire surface area with the outlines, including the required overlaps. The single step version works the same as the MIDDLE instruction, where the lines are shown for the next 0.6 meters from the current cleaning point.

4 USER STUDY

We evaluate the implemented pattern guidance instructions in a user study designed to test their efficacy for full surface coverage. Since the ideal pattern of surface coverage was already pre-established by the interested cleaning companies, we chose to keep this pattern within our study. We also focus our study on the efficacy of the patterns for full surface coverage. We omitted the cleanroom cleaning errors from the implementation to avoid complicating the procedure and influencing the pattern following results. This includes making the heatmap red when cleaning too fast, going in the wrong direction, or walking on top of the cleaned surface. These errors would most likely cause users to stop their current motion to cover these areas again, which causes bias in the results on whether they follow the pattern correctly with the visualizations. We did keep the rendering of the cleaned areas and overlaps since these are our core concepts to test for the assistance of the cleaning instructions included with the pattern guidance.

4.1 Hypotheses

H_1 *By providing instructions as a single step, pattern understandability improves.*

One of the results we want to measure in our study is the efficacy of providing instructions as a single step for surface coverage. In prior research of short singular motions, single steps have already been shown to be easier to understand due to less clutter (information you do not currently need yet) [16]. We pose H_1 to verify these claims and translate them over to surface coverage activities.

H_2 *Less surface coverage is achieved with instructions that do not show where the edges of the mop should be.*

In the design of our instructions, we already mentioned the expected limitation of the MIDDLE instruction (see subsubsection 3.4.3) in terms of communicating the exact orientation and position of the cleaning mop. To verify these claims, we test with H_2 whether other instructions that communicate the full position and orientation of the edges of the cleaning mop are more efficient at covering the entire surface.

H_3 *Instruction adherence is optimized by instructions that do not overlap with themselves.*

Since we also need to ensure overlaps within our surface coverage use case, we had to design the instructions to communicate these overlaps as well. For this reason, some of the instructions (BREADCRUMBS-FULL, OUTLINES-FULL) overlap with themselves, which we expect can cause confusion for the users on the correct part of the instruction to follow, hence we test for H_3.

H_4 *Instructions that cause users to look closer at the cleaning mop, achieve better surface coverage results.*

Instructions should not distract users from the real activity to ensure they carry it out correctly. We anticipate that some of the instructions cause users to look away from the cleaning activity. Looking away does not necessarily imply a bad execution of the activity, but we anticipate that an effect might persist in this behavior, hence we verify this by testing H_4.

4.2 Study Procedure

For the procedure of our study, we conducted a large-scale between-subjects design study where every participant only had one chance to cover the entire surface area with the assistance of the visualization assigned to them. Before conducting the experiment, they would be informed what the ideal cleaning pattern is, why it is important to follow such a pattern, and how the cleaning guidance system of the heatmap works (green coverage and overlaps). We also instruct them that during cleaning, they will need to replicate the pattern as accurately as possible and focus on covering the entire surface area, rather than being as fast as possible. Afterward, they are given the headset, which shows a starting point (footprint) at the corner of the canvas where they must stand. Once they stand on the footprint and face the correct direction, a gray rectangle

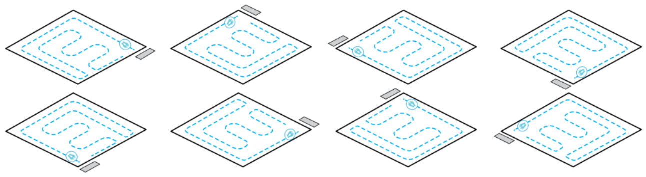


Figure 5: The configurations tested, the rectangle shows where the mop starts and the footprint where the user starts.

appears in front of them to indicate where they need to place the cleaning mop to start cleaning (at this point, this is outside the surface). These pre-study instructions can be seen on Figure 5, which also shows the variations of the pattern and the starting points we use. These variations are pseudo-randomized across the entire study, where we ensure all variations are present for all visualizations. Once the cleaning mop is placed correctly, the pattern visualization loads and show the user how they need to clean the surface. After they have finished cleaning the entire surface, they are prompted to fill in a final questionnaire. In this questionnaire, we first ask about their cleaning expertise. Afterward, we asked on a Likert-scale questionnaire (1-5) how understandable the pattern visualization was, how easy it was to follow the pattern, and how well the surface was fully cleaned. Lastly, we ask the participants whether they see uses for a tool like this in their daily lives.

4.3 Participants

For our study, we collaborated with a biotech company that was not part of the original design phase of this research. We conducted trials with 864 employees (365 male, 499 female) of varying backgrounds and ages (M=40.687, SD=11.58). In total, 62 had no cleaning experience, 367 cleaned sometimes at home, 409 participants cleaned often at home or other locations, and 25 cleaned professionally. We had to omit 15 trials from the performance measures due to unreliable tracking measurements caused by controller tracking malfunctions. The distribution of the participants for every visualization can be seen on Table 1.

Table 1: Amount of participants for every visualization.

	BREADCRUMBS	EXAMPLE	MIDDLE	OUTLINES
FULL	103	105	104	102
STEP	108	110	106	111

4.4 Data Collection

We take several measurements during the study to attempt to answer our hypotheses. For every frame of the application, we tracked the position and orientation of the mop, where the instructions are visible, and a ray cast of where the eyes are looking (using the built-in eye tracker from the Magic Leap 2). All the data is set to be relative to the platform to clean, and we automatically align the results according to the given pattern (from Figure 5). We also saved snapshots of the state of the heatmap (cleaned surface) after every participant finished cleaning in the form of a 512x512 image.

We used the think-aloud protocol during the study and asked several questions related to their performance after the activity (see subsection 4.2) to measure the usability of the instruction.

5 RESULTS

Since we conducted a between-subject design study, we performed the following tests to verify the significance of the results: Shapiro-Wilkinson to test if the data follows a normal distribution. If normal distribution can be assumed and we find that homogeneity of variances (through Bartlett's test) can be assumed, we perform ANOVA. Otherwise, we use Kruskal-Wallis to test for significance. We use the Pairwise Wilcoxon rank sum test with Holm correction as post-hoc test and calculate the individual effect sizes.

Since participants were asked to focus on accuracy rather than speed, the completion time became more varied (M=52.37 seconds, SD=19.57 seconds) across participants. However, since our hypotheses are not directly related to completion time, we do not discuss these results in depth.

5.1 Pattern Replication

To analyze whether users could follow the pattern correctly with the guidance visualizations, we use a technique called "Dynamic Time Warping" (DTW) [3]. This technique is used to calculate how far patterns are apart from each other, not by time, but by trying to match their sequences. Since we do not want the DTW distance to be based on the amount of data points (participants that take longer have more data points, thus a higher distance), we compare the visualizations based on the normalized DTW distance (distance divided by the amount of reference and query points). We compare the pattern of every participant to a singular baseline pattern (also used to render the instructions). The results of the pattern replication across the visualizations are shown on the top of Figure 6 (DTW Distance). Here we notice that the STEP versions of the instructions generally have lower DTW distances than the FULL instructions, with the largest gap in BREADCRUMBS and EXAMPLE. When testing the DTW distances for normality, we found they did not follow a normal distribution and found a significant difference ($\tilde{\chi}^2$=80.62, df=7, p<.001) using Kruskal-Wallis. The post-hoc test revealed a significant difference between BREADCRUMBS-FULL and EXAMPLE-FULL with all other visualizations (not with each other). We have listed the full results of the significance, and effect sizes, on Table 2 in Appendix A. Here we conclude that pattern replication was the worst for the BREADCRUMBS-FULL and EXAMPLE-FULL.

5.2 Cleaning Coverage

When comparing the cleaning results of the visualizations, we primarily look at how covered the surface was. Figure 7 shows the aggregated cleaning performance across all participants which shows how well the surfaces were covered. It is noticeable how the worst performance is achieved around the corners, especially for the FULL conditions and OUTLINES-STEP. To analyze the cleaning coverage statistically, we calculated how many pixels of the surface were cleaned (percentage-wise, where 100% implies the full surface is cleaned and 0% nothing is cleaned). The results of the percentages across the visualizations can be found on Figure 6 (Surface Coverage), where we see generally higher results for the

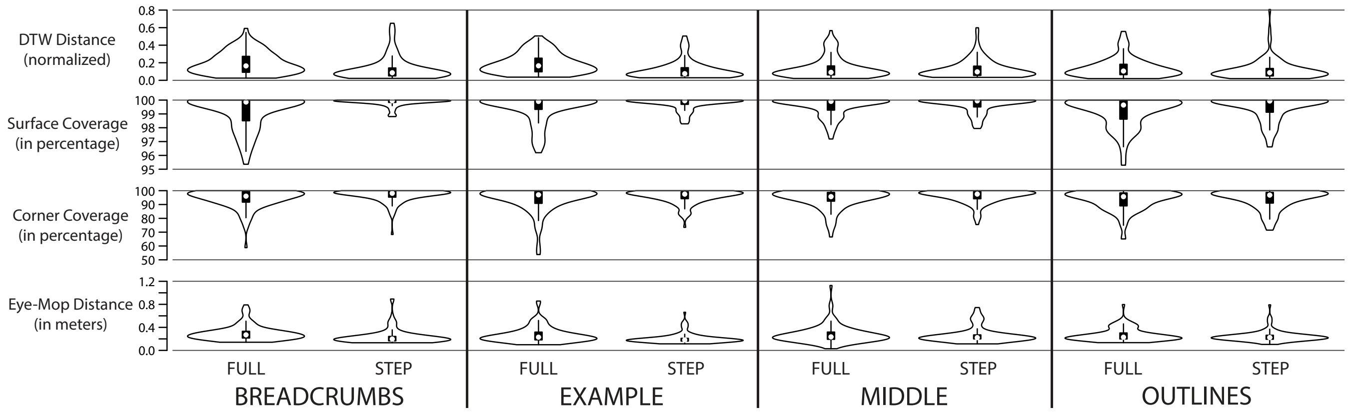


Figure 6: Violin plots of the full measurements of the performance during the activity.

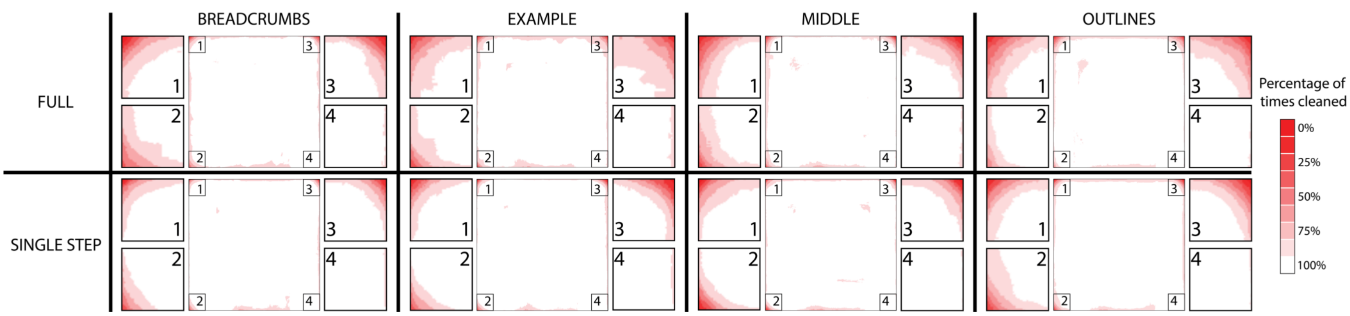


Figure 7: Coverage of the surface for every visualization across all participants, the colors are calculated based on every cleaned pixel, the corners we analyze separately are indicated by the numbered black squares and are upscaled for clarity

STEP conditions, in particular for the BREADCRUMBS-STEP and EXAMPLE-STEP conditions. We found the cleaning coverage to be non-normally distributed and found a significant difference between the visualizations ($\tilde{\chi}^2$=18.88, df=7, p=.0086). The post-hoc test reveals a significance between the OUTLINES-FULL and the BREADCRUMBS-STEP (p=.027, Z= 3.30, r=.23), EXAMPLE-STEP (p=.014, Z=3.48, r=.24), and MIDDLE-STEP (p=.037, Z=3.19, r=.22). From these results, we can state that OUTLINES-FULL had a worse cleaning coverage compared to the STEP conditions of the other visualizations. Even though we found a significant difference between some visualizations, in general, coverage of the cleaning activity across the entire surface was quite high (scale of the graph was between 95-100%). This is because small misses are not represented when scaling it up to a full surface.

To gain an additional understanding of the surface coverage performance, we analyze the results of the corners of the platform (1/8th of the total surface size) where most cleaning errors usually take place. We combine the results of the four corners for every participant together to gather the results listed on Figure 6 (Corner Coverage). Here we see again a better performance for the STEP versions of the visualizations. When testing the results for normality, we found that it does not follow a normal distribution and that there is a statistical significance ($\tilde{\chi}^2$=24.01, df=7,p=.00113). The post-hoc test then revealed a significance between BREADCRUMBS-STEP and MIDDLE-FULL (p=.043, Z=3.16, r=.22), OUTLINES-FULL (p=.0017, Z=4.02, r=.28), and OUTLINES-STEP (p=.05, Z=3.10, r=.21). From these results, we can conclude that corners were covered most accurately by the BREADCRUMBS-STEP visualization, where the second most accurate are EXAMPLE-STEP and MIDDLE-STEP.

5.3 Eye Tracking

To test how actively users look at the mop, we compute the eye-mop distance by raycasting the gaze from the camera and intersecting it with the height of the cleaning platform. We then take the Euclidean distance across the X-Z axes from the center of the mop to the gaze intersection to get an idea of where participants are looking on the surface. For every participant, we take the median eye-mop distance from the entire activity to get one value to represent the measures for that participant. The results of this measure can be seen on Figure 6 (Eye-Mop Distance). Again, we notice how the values are lower for the STEP conditions compared to the FULL conditions. In particular, the BREADCRUMBS-STEP and EXAMPLE-STEP report the lowest eye-mop distance compared to all other visualizations.

We found that the eye-mop data does not follow a normal distribution. When performing the Kruskal-Wallis test, we found a significant difference ($\tilde{\chi}^2$=82.77, df=7, p<.001) across the full results, whereas the post-hoc test revealed a significant difference between the BREADCRUMBS-FULL and all STEP visualizations. We also found significance for BREADCRUMBS-STEP and EXAMPLE-STEP with the MIDDLE and OUTLINES visualizations and between

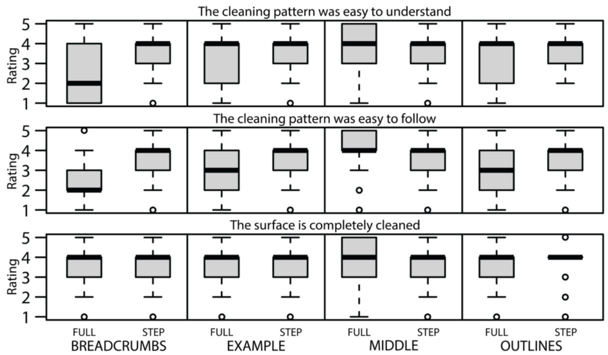


Figure 8: Likert-Scale Questionnaire ratings for the visualizations (1 Absolutely Disagree, 5 Absolutely Agree).

EXAMPLE-STEP and EXAMPLE-FULL (all details can be found on Table 3 in Appendix A). Combining these significance tests with the results we found for the eye-mop distance, the EXAMPLE-STEP visualization outperforms all others in terms of eye proximity to the mop, followed by BREADCRUMBS-STEP, then all other STEP conditions, and then finally the FULL conditions.

5.4 Questionnaire

Of the 863 participants (before performance removal), 132 found the tool useful for use at home, 293 found the tool useful for work contexts, and 438 did not see the need for such a tool. Those with an interest in the tool at home were particularly fond of seeing where the user had cleaned but saw no need for the cleaning pattern. We evaluate three statements related to the cleaning activity that were asked to participants in the post-study questionnaire in the form of a Likert-Scale of 1-5 (Absolutely Disagree-Absolutely Agree): "The cleaning pattern was easy to understand", "The cleaning pattern was easy to follow", and "The surface is completely cleaned". The results of these Likert-Scale questions are shown on Figure 8. For the ratings of these statements, it is noticeable how MIDDLE-FULL continuously had the highest reported ratings. For the first two statements, the STEP versions of the instructions had the second best ratings, with very similar results to each other, while the other FULL instructions had the lowest ratings. Finally, for the final statement "The surface is completely cleaned", all visualizations had similar results, except for the MIDDLE-FULL instruction. When we checked the results for normality, we found that none of the statement ratings followed a normal distribution. We will cover the results of each of these statement ratings more in-depth.

5.4.1 The cleaning pattern was easy to understand. Here, we found a significant result ($\tilde{\chi}^2$=85.85, df=7, p<.001), and the post-hoc test revealed a significance between the results of BREADCRUMBS-FULL and all other visualizations. We also found a significant difference between OUTLINES-FULL and BREADCRUMBS-STEP, MIDDLE-FULL, MIDDLE-STEP, and OUTLINES-STEP. Finally, we found a significant difference between MIDDLE-FULL and EXAMPLE-FULL. Full statistical test results are listed on Table 4 in Appendix B. Based on these results, we can testify that the BREADCCRUMBS-FULL made the pattern the hardest to understand from all the visualizations, and OUTLINES-FULL was the second hardest to understand. In contrast. the MIDDLE-FULL instruction was the easiest to understand from all the visualizations, followed by the STEP versions of the instructions (where no significance was found between them and the MIDDLE-FULL).

5.4.2 The cleaning pattern was easy to follow. For this statement, we found a significant result ($\tilde{\chi}^2$=96.47, df=7, p<.001) where the post-hoc test revealed statistical significances between BREADCRUMBS-FULL and all other visualizations (except for EXAMPLE-FULL) and a significance between EXAMPLE-FULL and the other visualizations (except OUTLINES-FULL). We also found a significance between OUTLINES-FULL and OUTLINES-STEP and found a significance between MIDDLE-FULL and all other visualizations (except MIDDLE-STEP and OUTLINES-STEP), where MIDDLE-FULL outperformed these visualizations. The full statistical values can be found on Table 5 in Appendix B. From these results, we found the worst ratings for the BREADCRUMBS-FULL condition, with the second-worst ratings coming from EXAMPLE-FULL, with OUTLINES-FULL following afterward. MIDDLE-FULL was rated the easiest to follow, with all the STEP conditions following after (EXAMPLE-STEP last).

5.4.3 The surface is completely cleaned. For this statement, we found an overall significant difference ($\tilde{\chi}^2$=17.42, df=7, p=.01487) and the post-hoc test revealed a significance between BREADCRUMBS-FULL and OUTLINES-STEP (p=.05, Z=3.12, r=.22). While the MIDDLE-FULL seems to have had the highest-rated performance (according to Figure 8), we can only reliably conclude from our findings that OUTLINES-STEP outperforms BREADCRUMBS-FULL in terms of perception of how covered the surface is.

6 DISCUSSION

Within our large-scale study, we found several significant findings that allow us to reason on our hypotheses. First and foremost, the STEP instructions continuously performed better than the FULL instructions for all the measures we tested (not always significant). In particular, the worst pattern and cleaning performance was achieved with the BREADCRUMBS-FULL condition and EXAMPLE-FULL second-worst. The best performance for the surface coverage and cleaning coverage was achieved with BREADCRUMBS-STEP condition. However, for the questionnaires, the MIDDLE-FULL condition had the highest ratings, followed by all the STEP conditions (better than the FULL counterparts). We expect that the perceived usability of the MIDDLE-FULL was because it was the clearest method to see the entire surface pattern before the activity. This did not guarantee better execution for the activity measures, often being outperformed by other visualizations (BREADCRUMBS-STEP significantly for corner coverage). From our results, we can safely accept H_1 stating that "By providing instructions as a single step, pattern understandability improves.", highlighting the positive influence of presenting instructions as a single step for motion guidance, similar to what was seen for other types of activities [12, 21, 28].

For the coverage of the surface area and the corners, we only found significant differences indicating that the OUTLINE-FULL was worse for full surface coverage and that BREADCRUMBS-STEP was better at getting users to clean the corners of the activity (significantly compared to MIDDLE-FULL). In other words, we could find no supporting evidence for H_2 stating that "Less surface coverage is achieved with instructions that do not show where the edges of the mop should be", which primarily questions the efficiency of the MIDDLE instruction for surface coverage. Hence we reject hypothesis H_2. This finding is important for the design of future instructions aimed at getting users to follow a pattern while still covering the outer edges of the activity, specifically since paths without edges such as MIDDLE are easy to understand (see questionnaire results on subsection 5.4) and are often already present within other guidance research [17, 18, 33, 34]. However, it is important to note that better results will most likely be achieved with other instructions such as BREADCRUMBS-STEP, especially for activities that require the pattern to cover large surface areas.

Another hypothesis we explored within our study is H_3 "Instruction adherence is optimized by instructions that do not overlap with themselves", where overlapping visualizations are primarily the BREADCRUMBS-FULL and OUTLINES-FULL. For surface and corner coverage, OUTLINES-FULL performed the worst, while for pattern replication and usability, BREADCRUMBS-FULL performed the worst. The STEP conditions, which have by default no overlap, also continuously performed better. Based on these results, we accept hypothesis H_3. In other words, when overlap is required during activities, instructions should not incorporate overlaps themselves since it can confuse users which part of the instruction needs to be followed at the time of movement. Providing information as a single step for such cases is a good alternative. The information to force users towards overlap is not lost, and usability improvements have been found (as shown by BREADCRUMBS-STEP).

The final hypothesis we defined within our study is H_4, which states, "Instructions that cause users to look closer at the cleaning mop, achieve better surface coverage results". For the distance of the gaze intersection to the mop, we found that users looked closest at the mop with the EXAMPLE-STEP and BREADCRUMBS-STEP instructions (in that order) and generally with the STEP instructions. We also found EXAMPLE-STEP and BREADCRUMBS-STEP to have the highest reported surface and corner coverage. Due to these performance increases, we accept the hypothesis H_4. From these findings, we highlight the importance of not using instructions that take the user's gaze away from the motion activity they are performing since this can directly affect the efficacy of the motion they perform, which is in line with previous observations [6, 22].

7 FUTURE WORK

Within this work, we have presented an AR cleaning system that can be used to highlight areas that were cleaned and how to guide cleaning patterns. Our participants were of very mixed backgrounds, with most having no professional cleaning experience. Of these participants, there were mixed responses towards the idea of using the technology at home or in work situations. This shows that, while skepticism remains, the interest in the adoption of the technology for such use cases is also present. In future work, it would be interesting to explore how the cleanroom operators feel about the technology and whether they think it can benefit their daily lives. While we saw performance increases with the STEP instructions, we did not test how much information the STEP instructions should present to achieve the ideal results. Future work should still explore how far ahead STEP instructions should be to keep distraction limited but also increase perceived usability (as shown by MIDDLE-FULL). We believe that, while we only covered a cleaning use case here, our findings can aid in the design of other surface coverage use cases (e.g., painting, plastering, vacuuming). Verifying these claims for other use cases in future research would be beneficial in expanding the number of tasks that can be supported with AR. In our study, we only analyzed measures related to the performance of pattern replication and surface coverage and omitted the errors that could occur (walked over cleaned surface, cleaned too fast, cleaned in the wrong direction). When errors occur, ideally the pattern guidance visualization should adapt to suggest the new ideal pattern to follow. However, to avoid error bias in our study, we did not study these adaptations. We also did not study yet how to combine other metrics such as communicating the correct speed with the visualizations. Future research should consider how to combine other metrics with the pattern guidance visualizations, including how error handling should occur, to optimize user guidance.

8 CONCLUSION

We have implemented an augmented reality cleaning guidance system focused on the typical processes within cleanroom cleaning to help support cleaning operators. We have conducted a large-scale study (n=864) using the guidance system, where we tried to understand what the most efficient way is to communicate the ideal cleaning pattern during the activity. We have found that presenting instructions as a single step proves to be beneficial in terms of optimizing pattern replication, cleaning coverage, activity focus, and usability. While providing a line in the middle of the activity is sufficient to achieve proper results, the best coverage is achieved by only showing the next step in the form of a static example (breadcrumb) where the user needs to move the cleaning mop. Instructions that overlap with themselves should be avoided, as these have a negative impact on the understanding of the task at hand. Presenting the instructions as a single step also allows for improvements in eye focus toward the cleaning mop, which impacts how the motion activity is conducted. While our cleaning guidance system proved to be effective in achieving full cleaning coverage results, we expect our findings related to the pattern guidance to be applicable for other pattern-following or surface coverage use cases other than just cleanroom cleaning.

ACKNOWLEDGMENTS

This work was made possible with support from the projects XR-Huis and MAXVR-INFRA subsidized by the Flemish Government and the European Union. We also want to thank the two cleanroom cleaning companies for providing design input and the employees from the study company for their participation in the study.

REFERENCES

[1] Salvador S. Agati, Rudieri D. Bauer, Marcelo Da S. Hounsell, and Aleksander S. Paterno. 2020. Augmented Reality for Manual Assembly in Industry 4.0: Gathering Guidelines. In *2020 22nd Symposium on Virtual and Augmented Reality (SVR)*. IEEE, 179–188. https://doi.org/10.1109/SVR51698.2020.00039 Place: Porto de Galinhas, Brazil.

[2] Per Backlund, Henrik Engstrom, Cecilia Hammar, Mikael Johannesson, and Mikael Lebram. 2007. Sidh - a Game Based Firefighter Training Simulation. In *2007 11th International Conference Information Visualization (IV '07)*. IEEE, 899–907. https://doi.org/10.1109/IV.2007.100 Place: Zurich, Switzerland.

[3] Donald J. Berndt and James Clifford. 1994. Using dynamic time warping to find patterns in time series. In *Proceedings of the 3rd International Conference on Knowledge Discovery and Data Mining (AAAIWS'94)*. AAAI Press, 359–370.

[4] Jeroen Ceyssens, Bram Van Deurzen, Gustavo Rovelo Ruiz, Kris Luyten, and Fabian Di Fiore. 2023. AR Guidance Design for Line Tracing Speed Control. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 1055–1063. https://doi.org/10.1109/ISMAR59233.2023.00122 Place: Sydney, Australia.

[5] Lea M. Daling and Sabine J. Schlittmeier. 2024. Effects of Augmented Reality-, Virtual Reality-, and Mixed Reality–Based Training on Objective Performance Measures and Subjective Evaluations in Manual Assembly Tasks: A Scoping Review. *Human Factors: The Journal of the Human Factors and Ergonomics Society* 66, 2 (Feb. 2024), 589–626. https://doi.org/10.1177/00187208221105135

[6] Hans J. Eysenck and Warren Thompson. 1966. THE EFFECTS OF DISTRACTION ON PURSUIT ROTOR LEARNING, PERFORMANCE AND REMINISCENCE. *British Journal of Psychology* 57, 1 (May 1966), 99–106. https://doi.org/10.1111/j.2044-8295.1966.tb01009.x

[7] Mana Fukasawa and Yu Nakayama. 2022. Spatial Augmented Reality Assistance System with Accelerometer and Projection Mapping at Cleaning Activities. In *ACM SIGGRAPH 2022 Posters*. ACM, 1–2. https://doi.org/10.1145/3532719.3543245 Place: Vancouver BC Canada.

[8] S. Garrido-Jurado, R. Muñoz-Salinas, F.J. Madrid-Cuevas, and M.J. Marín-Jiménez. 2014. Automatic generation and detection of highly reliable fiducial markers under occlusion. *Pattern Recognition* 47, 6 (June 2014), 2280–2292. https://doi.org/10.1016/j.patcog.2014.01.005

[9] Janet K. Gibbs, Marco Gillies, and Xueni Pan. 2022. A comparison of the effects of haptic and visual feedback on presence in virtual reality. *International Journal of Human-Computer Studies* 157 (Jan. 2022), 102717. https://doi.org/10.1016/j.ijhcs.2021.102717

[10] Florian Jeanne, Indira Thouvenin, and Alban Lenglet. 2017. A study on improving performance in gesture training through visual guidance based on learners' errors. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*. ACM, 1–10. https://doi.org/10.1145/3139131.3139144 Place: Gothenburg Sweden.

[11] Lasse Jensen and Flemming Konradsen. 2018. A review of the use of virtual reality head-mounted displays in education and training. *Education and Information Technologies* 23, 4 (July 2018), 1515–1529. https://doi.org/10.1007/s10639-017-9676-0

[12] Krithikashree Lakshminarayanan, Mayank Arvindbhai Patel, Zia Din, and Lingguang Song. 2023. Lessons Learned from Developing and Testing an Augmented Reality Application for Just-in-Time Information Delivery to Improve Construction Safety. In *Proceedings of the Canadian Society of Civil Engineering Annual Conference 2022*, Rishi Gupta, Min Sun, Svetlana Brzev, M. Shahria Alam, Kelvin Tsun Wai Ng, Jianbing Li, Ashraf El Damatty, and Clark Lim (Eds.). Vol. 363. Springer International Publishing, Cham, 229–245. https://doi.org/10.1007/978-3-031-34593-7_15

[13] Benedikt Lauber and Martin Keller. 2014. Improving motor performance: Selected aspects of augmented feedback in exercise and health. *European Journal of Sport Science* 14, 1 (Feb. 2014), 36–43. https://doi.org/10.1080/17461391.2012.725104

[14] Dieter Lerner, Stefan Mohr, Jonas Schild, Martin Göring, and Thomas Luiz. 2020. An Immersive Multi-User Virtual Reality for Emergency Simulation Training: Usability Study. *JMIR Serious Games* 8, 3 (July 2020), e18822. https://doi.org/10.2196/18822

[15] Jen-Shuo Liu, Carmine Elvezio, Barbara Tversky, and Steven Feiner. 2021. Using Multi-Level Precueing to Improve Performance in Path-Following Tasks in Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* 27, 11 (Nov. 2021), 4311–4320. https://doi.org/10.1109/TVCG.2021.3106476

[16] Jen-Shuo Liu, Barbara Tversky, and Steven Feiner. 2022. Precueing Sequential Rotation Tasks in Augmented Reality. In *Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology*. ACM, 1–11. https://doi.org/10.1145/3562939.3565641 Place: Tsukuba Japan.

[17] Jen-Shuo Liu, Barbara Tversky, and Steven Feiner. 2023. Cueing Sequential 6DoF Rigid-Body Transformations in Augmented Reality. In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 356–365. https://doi.org/10.1109/ISMAR59233.2023.00050 Place: Sydney, Australia.

[18] Jen-Shuo Liu, Portia Wang, Barbara Tversky, and Steven Feiner. 2022. Adaptive Visual Cues for Guiding a Bimanual Unordered Task in Virtual Reality. In *2022 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 431–440. https://doi.org/10.1109/ISMAR55827.2022.00059 Place: Singapore, Singapore.

[19] Masashi Narita and Takafumi Matsumaru. 2015. Calligraphy-stroke learning support system using projection. In *2015 24th IEEE International Symposium on Robot and Human Interactive Communication (RO-MAN)*. IEEE, 640–645. https://doi.org/10.1109/ROMAN.2015.7333576 Place: Kobe, Japan.

[20] Akira Nomoto, Yuki Ban, Takuji Narumi, Tomohiro Tanikawa, and Michitaka Hirose. 2016. Supporting Precise Manual-handling Task using Visuo-haptic Interaction. In *Proceedings of the 7th Augmented Human International Conference 2016*. ACM, 1–8. https://doi.org/10.1145/2875194.2875216 Place: Geneva Switzerland.

[21] Junghwan Park, Meelim Kim, Mohamed El Mistiri, Rachael Kha, Sarasij Banerjee, Lisa Gotzian, Guillaume Chevance, Daniel E Rivera, Predrag Klasnja, and Eric Hekler. 2023. Advancing Understanding of Just-in-Time States for Supporting Physical Activity (Project JustWalk JITAI): Protocol for a System ID Study of Just-in-Time Adaptive Interventions. *JMIR Research Protocols* 12 (Sept. 2023), e52161. https://doi.org/10.2196/52161

[22] Jay Pratt and Richard A. Abrams. 1994. Action-centered inhibition: Effects of distractors on movement planning and execution. *Human Movement Science* 13, 2 (June 1994), 245–254. https://doi.org/10.1016/0167-9457(94)90039-6

[23] Aylen Ricca, Amine Chellali, and Samir Otmane. 2020. Influence of hand visualization on tool-based motor skills training in an immersive VR simulator. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 260–268. https://doi.org/10.1109/ISMAR50242.2020.00049 Place: Porto de Galinhas, Brazil.

[24] Tim Sandle. 2012. *The CDC Handbook: A Guide to Cleaning and Disinfecting Cleanrooms*.

[25] Maurice Suckling. 2020. If It Works, Break It. In *Professional Techniques for Video Game Writing* (2 ed.), Wendy Despain (Ed.). CRC Press, 257–269. https://doi.org/10.1201/9780429196539-19

[26] Spyros Vosinakis and Panayiotis Koutsabasis. 2018. Evaluation of visual feedback techniques for virtual grasping with bare hands using Leap Motion and Oculus Rift. *Virtual Reality* 22, 1 (March 2018), 47–62. https://doi.org/10.1007/s10055-017-0313-4

[27] Zhuo Wang, Xiaoliang Bai, Shusheng Zhang, Mark Billinghurst, Weiping He, Peng Wang, Weiqi Lan, Haitao Min, and Yu Chen. 2022. A comprehensive review of augmented reality-based instruction in manual assembly, training and repair. *Robotics and Computer-Integrated Manufacturing* 78 (Dec. 2022), 102407. https://doi.org/10.1016/j.rcim.2022.102407

[28] Nicholas C. Wilson, Mehmet Kosa, Casper Harteveld, Mohsen Moghaddam, and Kemi Jona. 2023. Exploring Extended Reality for Scaffolding Psychomotor-Intensive Tasks in Complex Real-World Settings. 2077–2078. https://doi.org/10.22318/icls2023.177665

[29] Biao Xie, Huimin Liu, Rawan Alghofaili, Yongqi Zhang, Yeling Jiang, Flavio Destri Lobo, Changyang Li, Wanwan Li, Haikun Huang, Mesut Akdere, Christos Mousas, and Lap-Fai Yu. 2021. A Review on Virtual Reality Skill Training Applications. *Frontiers in Virtual Reality* 2 (April 2021), 645153. https://doi.org/10.3389/frvir.2021.645153

[30] Ungyeon Yang and Gerard Jounghyun Kim. 2002. Implementation and Evaluation of "Just Follow Me": An Immersive, VR-Based, Motion-Training System. *Presence: Teleoperators and Virtual Environments* 11, 3 (June 2002), 304–323. https://doi.org/10.1162/105474602317473240

[31] Xin Yi, Xueyang Wang, Jiaqi Li, and Hewu Li. 2023. Examining the Fine Motor Control Ability of Linear Hand Movement in Virtual Reality. In *2023 IEEE Conference Virtual Reality and 3D User Interfaces (VR)*. IEEE, 427–437. https://doi.org/10.1109/VR55154.2023.00058 Place: Shanghai, China.

[32] Kevin Yu, Daniel Ostler, Jonas Fuchtmann, Anna Zapaishchykova, Maximilian Berlet, Nassir Navab, Hubertus Feussner, and Dirk Wilhelm. 2021. Clean-AR: Using Augmented Reality for Reducing the Risk of Contamination from Airborne Disease Agents on Surfaces. *Current Directions in Biomedical Engineering* 7, 1 (Aug. 2021), 6–10. https://doi.org/10.1515/cdbme-2021-1002

[33] Xingyao Yu. 2023. [DC] Limb Motion Guidance in Extended Reality. In *2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*. IEEE, 967–968. https://doi.org/10.1109/VRW58643.2023.00326 Place: Shanghai, China.

[34] Xingyao Yu, Katrin Angerbauer, Peter Mohr, Denis Kalkofen, and Michael Sedlmair. 2020. Perspective Matters: Design Implications for Motion Guidance in Mixed Reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 577–587. https://doi.org/10.1109/ISMAR50242.2020.00085 Place: Porto de Galinhas, Brazil.

[35] Shang Zhao, Xiao Xiao, Qiyue Wang, Xiaoke Zhang, Wei Li, Lamia Soghier, and James Hahn. 2020. An Intelligent Augmented Reality Training Framework for Neonatal Endotracheal Intubation. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. IEEE, 672–681. https://doi.org/10.1109/ISMAR50242.2020.00097 Place: Porto de Galinhas, Brazil.

A REPORTED STATISTICAL SIGNIFICANCES OF MEASUREMENTS

Table 2: Dynamic Time Warp normalized distance P-values of the pairwise Wilcoxon rank sum test results, bold highlights significance and parentheses give the Z values and effect sizes (r) for the significant values.

	BREADCRUMBS-FULL	BREADCRUMBS-STEP	EXAMPLE-FULL	EXAMPLE-STEP	MIDDLE-FULL	MIDDLE-STEP	OUTLINES-FULL
BREADCRUMBS-STEP	**8.3e-06 (Z=5.09, r=0.35)**	-	-	-	-	-	-
EXAMPLE-FULL	1.00000	**3.0e-06 (Z=5.29, r=0.36)**	-	-	-	-	-
EXAMPLE-STEP	**3.9e-07 (Z=5.66, r=0.39)**	1.00000	**1.1e-07 (Z=5.88, r=0.40)**	-	-	-	-
MIDDLE-FULL	**0.00079 (Z=4.10, r=0.29)**	1.00000	**0.00028 (Z=4.35, r=0.30)**	1.00000	-	-	-
MIDDLE-STEP	**5.5e-05 (Z=4.70, r=0.33)**	1.00000	**3.2e-05 (Z=4.82, r=0.33)**	1.00000	1.00000	-	-
OUTLINES-FULL	**0.00456 (Z=3.65, r=0.26)**	1.00000	**0.00376 (Z=3.71, r=0.26)**	0.40412	1.00000	1.00000	-
OUTLINES-STEP	**2.2e-07 (Z=5.76, r=0.39)**	1.00000	**8.9e-08 (Z=5.92, r=0.40)**	1.00000	1.00000	1.00000	0.54815

Table 3: Eyes projection and mop distance P-values of the pairwise Wilcoxon rank sum test results, bold highlights significance and parentheses give the Z values and effect sizes (r) for the significance.

	BREADCRUMBS-FULL	BREADCRUMBS-STEP	EXAMPLE-FULL	EXAMPLE-STEP	MIDDLE-FULL	MIDDLE-STEP	OUTLINES-FULL
BREADCRUMBS-STEP	**1.1e-06 (Z=5.49, r=0.39)**	-	-	-	-	-	-
EXAMPLE-FULL	0.14062	0.16694	-	-	-	-	-
EXAMPLE-STEP	**3.5e-12 (Z=7.41, r=0.52)**	0.22787	**0.00011 (Z=4.56, r=0.32)**	-	-	-	-
MIDDLE-FULL	0.63067	**0.01522 (Z=3.35, r=0.24)**	1.00000	**3.0e-06 (Z=5.28, r=0.37)**	-	-	-
MIDDLE-STEP	**0.03156 (Z=3.12, r=0.22)**	**0.03156 (Z=3.13, r=0.22)**	1.00000	**2.4e-06 (Z=5.33, r=0.37)**	1.00000	-	-
OUTLINES-FULL	0.42716	**0.00267 (Z=3.83, r=0.28)**	1.00000	**4.6e-08 (Z=6.03, r=0.43)**	1.00000	1.00000	-
OUTLINES-STEP	**0.00382 (Z=3.73, r=0.26)**	**0.03167 (Z=3.09, r=0.22)**	1.00000	**1.6e-06 (Z=5.40, r=0.38)**	1.00000	1.00000	1.00000

B REPORTED STATISTICAL SIGNIFICANCES OF QUESTIONNAIRE RATINGS

Table 4: Likert-Scale Ratings of "The cleaning pattern was easy to understand" P-values of the pairwise Wilcoxon rank sum test results, bold highlights significance, and parentheses give the Z values and effect sizes (r) for the significant values.

	BREADCRUMBS-FULL	BREADCRUMBS-STEP	EXAMPLE-FULL	EXAMPLE-STEP	MIDDLE-FULL	MIDDLE-STEP	OUTLINES-FULL
BREADCRUMBS-STEP	**5.5e-09 (Z=6.35, r=0.46)**	-	-	-	-	-	-
EXAMPLE-FULL	**0.00016 (Z=4.49, r=0.33)**	0.29405	-	-	-	-	-
EXAMPLE-STEP	**1.1e-07 (Z=5.86, r=0.42)**	1.00000	1.00000	-	-	-	-
MIDDLE-FULL	**4.8e-10 (Z=6.73, r=0.50)**	1.00000	**0.02881 (Z=3.16, r=0.23)**	0.71993	-	-	-
MIDDLE-STEP	**3.1e-09 (Z=6.44, r=0.47)**	1.00000	0.15532	1.00000	1.00000	-	-
OUTLINES-FULL	**0.00184 (Z=3.92, r=0.29)**	**0.03544 (Z= 3.08, r=0.23)**	1.00000	0.20823	**0.00181 (Z=3.94, r=0.30)**	**0.01545 (Z=3.35, r=0.25)**	-
OUTLINES-STEP	**6.1e-10 (Z=6.69, r=0.48)**	1.00000	0.12272	1.00000	1.00000	1.00000	**0.00984 (Z=3.49, r=0.26)**

Table 5: Likert-Scale Ratings of "The cleaning pattern was easy to follow" P-values of the pairwise Wilcoxon rank sum test results, bold highlights significance, and parentheses give the Z values and effect sizes (r) for the significant values.

	BREADCRUMBS-FULL	BREADCRUMBS-STEP	EXAMPLE-FULL	EXAMPLE-STEP	MIDDLE-FULL	MIDDLE-STEP	OUTLINES-FULL
BREADCRUMBS-STEP	**1.8e-06 (Z=5.38, r=-0.39)**	-	-	-	-	-	-
EXAMPLE-FULL	0.48466	**0.03202 (Z=3.07, r=0.22)**	-	-	-	-	-
EXAMPLE-STEP	**2.4e-06 (Z=-5.32, r=0.38)**	1.00000	**0.04665 (Z=2.94, r=0.21)**	-	-	-	-
MIDDLE-FULL	**2.8e-12 (Z=7.44, r=0.55)**	**0.01524 (Z=3.32, r=0.24)**	**3.9e-07 (Z=5.66, r=0.42)**	**0.00901 (Z=3.48, r=0.25)**	-	-	-
MIDDLE-STEP	**5.7e-08 (Z=5.99, r=0.44)**	0.90943	**0.00206 (Z=3.87, r=0.28)**	0.88774	0.26313	-	-
OUTLINES-FULL	**0.00077 (Z=4.12, r=0.31)**	0.72977	0.72977	0.87939	**8.3e-05 (Z=4.62, r=0.35)**	0.15893	-
OUTLINES-STEP	**4.2e-10 (Z=6.74, r=0.49)**	0.72977	**0.00012 (Z=4.54, r=0.33)**	0.54308	0.72977	1.00000	**0.02033 (Z=3.22, r=0.24)**