# Cueing Sequential 6DoF Rigid-Body Transformations in Augmented Reality

Jen-Shuo Liu*
Department of Computer Science
Columbia University

Barbara Tversky†
Department of Human Development
Teachers College
Columbia University

Steven Feiner‡
Department of Computer Science
Columbia University

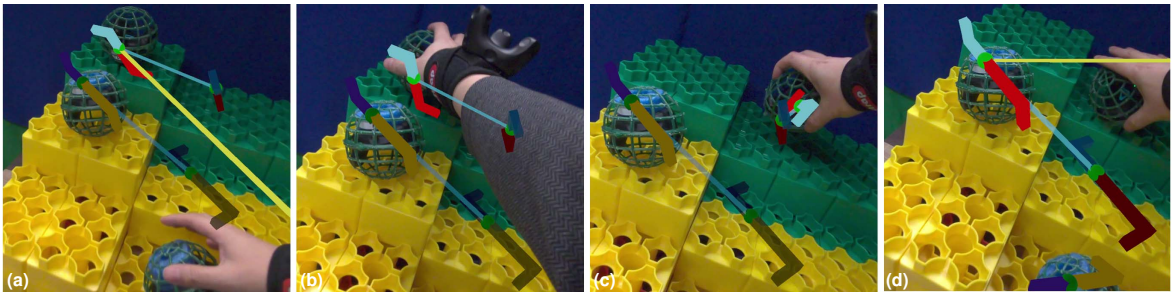


Figure 1: A step in our 6DoF task, viewed through a video–see-through AR headset. (a) The user is moving their dominant hand to a physical spherical object to be manipulated on the green block at the upper left, guided by a yellow line emanating from a wrist-worn tracker to that object. A cyan line connects the object to its destination at its right, while a pair of multi-colored "Z"s (each formed from a pair of "L"s and a small green sphere where they meet) show the rotation to be performed by aligning the "Z" attached to the object with the "Z" at the destination. A darker cyan line and a pair of brown-and-purple "Z"s document the next step, for another object on the upper yellow block. (b) The user picks up the current object and the yellow line has disappeared. (c) The user has transported the object following its cyan line and is completing its 3D rotation, guided by the "Z"s. (d) The user has deposited the object in the designated 6DoF pose. The visualizations are updated, and the user's hand is now connected to the next object.

## Abstract

Augmented reality (AR) has been used to guide users in multi-step tasks, providing information about the current step (*cueing*) or future steps (*precueing*). However, existing work exploring cueing and precueing a series of rigid-body transformations requiring rotation has only examined one-degree-of-freedom (DoF) rotations alone or in conjunction with 3DoF translations. In contrast, we address sequential tasks involving 3DoF rotations and 3DoF translations. We built a testbed to compare two types of visualizations for cueing and precueing steps. In each step, a user picks up an object, rotates it in 3D while translating it in 3D, and deposits it in a target 6DoF pose. Action-based visualizations show the actions needed to carry out a step and goal-based visualizations show the desired end state of a step. We conducted a user study to evaluate these visualizations and the efficacy of precueing. Participants performed better with goal-based visualizations than with action-based visualizations, and most effectively with goal-based visualizations aligned with the Euler axis. However, only a few of our participants benefited from precues, most likely because of the cognitive load of 3D rotations.

**Index Terms:** Human-centered computing—Human-centered computing (HCI)—Interaction paradigms—Mixed / augmented reality; Human-centered computing—Human-centered computing (HCI)—HCI design and evaluation methods—User studies

## 1 Introduction

Many real-life tasks involve a series of steps requiring users to pick up an object, translate it while rotating it, and place it in a designated six-degree-of-freedom (DoF) pose. For example, when assembling a piece of equipment, a user often needs to pick up one part at a time, translate it while rotating it, and place it at a target position and orientation. Similarly, when scanning an object with a hand-held camera, a user may need to sequentially move the camera to designated poses to take photos [2, 12, 43]. Augmented reality (AR) or virtual reality (VR) can be used to display task instructions to facilitate performance to help users who are unfamiliar with a task or for procedures that are input-dependent (e.g., [4, 18, 32, 37, 53]).

For sequential tasks with multiple steps, instructions can be classified as either cues or precues. A *cue* provides information about the current step, while a *precue* provides information about a future step. Previous studies have examined the effectiveness of precueing in VR and AR for improving user performance [19, 48]. Research on cueing and precueing sequential rigid-body transformations involving rotations has focused on 1DoF rotations alone [21] or in combination with 3DoF translations [20]. Other studies have investigated cueing for a single 3DoF rotation alone [41] or with a 3D translation [1]. However, for sequential 6DoF rigid-body transformations, it remains unclear which types of visualizations, and how many of them, would be most beneficial to users.

To address these issues, we developed an AR testbed that requires users to pick up a physical object in each step, rotate it in 3D while translating it in 3D, and place it in a target 6DoF pose (Figure 1). We designed two types of cueing and precueing visualizations, one based on the actions needed to achieve the goal (action-based) and the other based on the goal itself (goal-based), and assessed their effectiveness in a user study. We make the following contributions:

- We explore cueing and precueing sequential tasks involving 3DoF translations and 3DoF rotations.
- We develop goal-based and action-based visualizations, and show that participants performed better with our goal-based visualizations than with our action-based visualization.
- We show that aligning goal-based visualizations with the Euler axis improved task performance compared to a visualization that was not aligned. However, adding action information to a goal-based visualization did not improve performance.

*e-mail: jl5004@columbia.edu
†e-mail: bt2158@tc.columbia.edu
‡e-mail: feiner@cs.columbia.edu

2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR) | 979-8-3503-2838-7/23/$31.00 ©2023 IEEE | DOI: 10.1109/ISMAR59233.2023.00050



979-8-3503-2838-7/23/$31.00 ©2023 IEEE
DOI 10.1109/ISMAR59233.2023.00050

- We observe that only a few study participants benefited from a precue and discuss possible reasons based on previous work.

## 2 Related Work

### 2.1 Systems That Use Cueing and Precueing in AR/VR

Zhao et al. [52] explored the use of landmark cues in AR navigation. They found that world-fixed directions improved spatial learning but adding AR landmark cues only marginally improved spatial learning in a screen-fixed condition. Sodhi et al. [36] projected cues directly onto a user's hand to guide hand movement, but did not address rotation. These projects focused on cueing body or hand translations, but not object manipulation.

Polvi et al. [30] developed a handheld AR system to guide inspection tasks. An arrowhead cued translation, and an image of the desired viewpoint cued translation and rotation for a single step. The handheld AR device improved performance compared to a non-AR device. Shingu et al. [35] used AR to guide users to take a photo by moving their camera to a desired position and orientation. A sphere showed a point of interest and a cone originating from the sphere cued the user to move and orient the camera to a range of 6DoF poses. Hartl et al. [14] developed AR visualizations to cue users to locate a phone camera at a target 6DoF pose, with the best visualizations taking under 15 seconds. Sukan et al. [41], compared visualizations to guide a single 3DoF rotation of an object to a target orientation. They found that adding a pair of virtual "handles" to a task object to be aligned with a pair of target rings at the destination was more effective than a continuously computed circular arrow based on Euler's rotation theorem [8]. All these systems visualized only a single step at a time.

Andersen et al. [2] used an AR headset to guide users to take a series of photos with a handheld phone for scene reconstruction. Spheres linked by lines cued and precued 3DoF translations, while cones cued and precued 2DoF rotation (i.e., not including 1DoF roll about the cone axis). When the phone was close to its current destination, a 3D rectangle appeared, cueing the full 6DoF pose. However, their evaluation compared only the reconstruction process with and without this guidance, showing that it improved quality. Later, Andersen and Popescu [1] developed visualizations to guide a single 6DoF pose of a handheld controller. They tested rotations between 15° and 45°, and found that using a 3D coordinate-system gnomon centered at the object and another at the destination took the least time.

Schön et al. [34] examined mid-air rotation of virtual objects manipulated with a pinch gesture. They used a single visualization approach to explore how factors such as the orientation of the visible rotation axis and the object location relative to the user affected performance. In contrast, we examine combined 6DOF rotation and translation of physical objects, and explore the relative effectiveness of different visualization approaches for guiding such tasks.

### 2.2 Studies of AR/VR Precueing Efficacy

Volmer et al. [48] studied the use of a single movement precue in a projector-based spatial AR system and found that a line precue connecting the current and next destinations improved performance the most. In a further study using electroencephalography [49], they found that cognitive load could be reduced with proper precueing. Liu et al. [19] examined the effectiveness of using multiple translation precues in a VR path-following task, showing that participants could effectively use two to three precues. In subsequent studies in AR, when only 1DoF rotations were required, users were able to use rotation precues [21]. However, when translation was also included, user performance was largely determined by the translation precue [20, 21].

Unlike the previous work discussed above, we address cueing and precueing 6DoF sequential rigid-body transformations that include 3DoF translations and 3DoF rotations. Since lines have been shown to be useful for guiding translation [19, 48], we use lines to show translation, and focus on the design of visualizations to show rotation in sequential 6DoF rigid-body transformations. Inspired by Liu et al. [19], which showed that the best number of precues and visualization style depend on each other, we adopt a study design that tests both at the same time, presented in Section 5.

## 3 Multi-Step AR Task

We developed a compound task that involves a series of steps requiring the user to perform specific actions. In the spirit of Gilbreth and Gilbreth [10], we divide each step into distinct phases: At the start of each step, the user moves their unloaded hand to the designated object (Figure 1a) and picks it up (Figure 1b). The user then rotates the object in 3D while translating it in 3D (Figure 1c). Finally, they put the object down in a target 6DoF pose (Figure 1d), completing the step. Each step must be completed before the user can proceed to the next one. We determine if a step is complete by verifying whether the pose of the task object matches the designated 6DoF pose (within tolerances of 3cm and 15° of the designated pose). Therefore, we do not impose any constraints on how users translate and rotate the task object to achieve the desired pose.

Euler's rotation theorem [8] specifies that the transformation between any two orientations in 3D space can be accomplished by a single rotation around an optimal axis (Euler axis). We created the goal orientations for the steps by randomly pre-generating rotation axes and randomly pre-selecting from rotation angles of 70°, 100°, 130°, and 160°. In each step, the user was asked to pick up the object from one of the 15 target positions (Section 4.4), rotate the object by one of the four angles against a randomly generated axis while translating the object, and deposit it at another target position. We expected step difficulty would be determined by both the required translation and rotation.

For a condition with $m$ precued steps, when the user is performing step $k$, information about that particular step is cued, while details about future steps $k+1$, $k+2$, ..., $k+m$ are precued. After the user completes step $k$, step $k+1$ becomes the new cued step, and steps $k+2$, $k+3$, ..., $k+m+1$ become the new precued steps. This pattern continues until the user completes the entire task.

## 4 Visualizations

We initially considered a range of visualizations. One visualization we tried was a variation of the *Animate* visualization in Sukan et al. [41] that also included translation. However, we found it distracting and difficult to follow, particularly when the task object had a symmetrical shape and the rotation occurred outside of the user's current field of view. We also experimented with a visualization that used lines to connect points on the task object to their corresponding endpoints on the destination, inspired by Oda et al. [26]. However, these lines could be challenging to distinguish when they overlapped.

Noticing that these and other visualizations in previous work show either the action needed to complete a step or the goal of the step, we designed two types of visualizations to cue and precue translations and rotations: *action-based visualization* and *goal-based visualization*, described below. Both were designed to work independently of the shape of the task objects.

### 4.1 Action-based Visualization

We use "action-based" to indicate visualizations that show the required rotation action. Our action-based visualization shows how the hand should rotate and translate the object rather than showing the goal pose.

**Action-based-Euler (*AE*)**: AE (Figure 2) was inspired by the *SingleAxis* visualization of Sukan et al. [41], which uses Euler's rotation theorem [8]. However, AE visualizes 3DoF translation

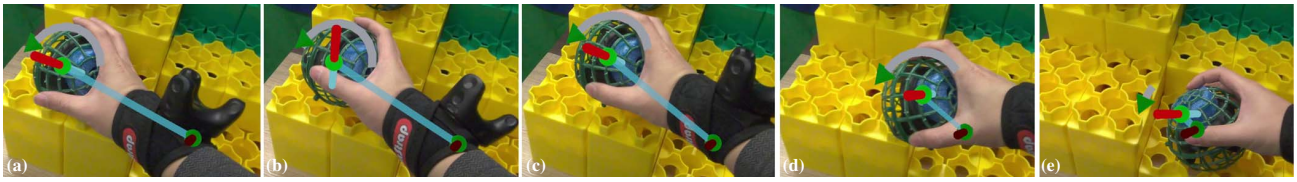


Figure 2: Action-based visualization. (a) A multi-colored cylinder with a small green sphere at its center shows the Euler axis on the task object, while a circular arrow indicates the suggested rotation. A darker multi-colored cylinder is at the step destination. A cyan line connects the two cylinders. (b) The user rotates the task object around a slightly different axis than the Euler axis. The multi-colored cylinder on the task object will rotate to reflect this and the user knows the cylinder should be rotated back. The cylinder at the step destination remains unchanged. (c) The user rotates the cylinder back to correct the rotation. (d–e) The user translates and rotates the object, and the tail of the circular arrow updates accordingly to show the new suggested rotation.

in addition to 3DoF rotation. To represent the required rotation for a step, we convert it to an angle–axis representation and use conjoined cyan and red cylinders superimposed on the task object to display the Euler axis. We chose to show a single rotation around the Euler axis rather than three rotations about the x-, y-, and z-axes because Sukan et al. [41] had shown the former to be more effective. A small green sphere is placed where the two cylinders meet to help the user recognize the center of the visualization. A circular arrow shows the required rotation angle in a step, following the convention of using arrows to depict rigid-body transformations in instruction manuals [25]. Another darker pair of conjoined cyan and red cylinders that also show the Euler axis are at the step destination.

We advised users to rotate the object to the designated pose with a rotation less than 180°. Research has shown that larger rotations are hard to perform with a single hand motion [42, 50] and therefore harder to mentally simulate [28, 29]. The tail of the arrow updates to reflect the amount of rotation around the Euler axis, while the arrowhead remains fixed relative to the rotation axis. To prevent the user's hand from colliding with the blocks, we limit the circular arrows from going below the object in the $-y$ direction.

When the user rotates the task object, the axis visualization on the task object is also rotated, but the axis visualization on the step destination remains unchanged. This is necessary when translating the task object because the user may rotate the object in a way that does not align with the Euler axis computed at the start of the step. While continuously recomputing the Euler axis is one solution used by Sukan et al. [41], they reported that this frustrated their participants, which we also noticed.

Instead, we compute the Euler axis only at the beginning of the step, while we continuously update the remaining rotation amount (angle around the Euler axis determined at the beginning of the step) shown by the rotation arrow. When the user rotates the task object around an axis other than the Euler axis, the conjoined cyan and red cylinders superimposed on the object will rotate to reflect this (Figure 2b) and no longer be parallel to the ones at the step destination, so the user knows the cylinder should be rotated back (Figure 2c).

### 4.2 Goal-based Visualizations

In contrast to our action-based visualization, our goal-based visualizations explicitly indicate the intended destination and orientation to guide users to compute the rotation without visualizing the action. The aim of goal-based visualizations is to place visualizations on the task object and the destination to make their orientations clear to the user. This helps the user compute the required rotation by themselves. In a previous iteration, we utilized a semi-transparent replica of the task object at the desired destination with its corresponding pose. To help users differentiate between the two identical shapes, we added a small shape to the replica. However, internal testing revealed that users still experienced confusion due to the symmetric shape of the object.

To overcome this confusion, we added a shape that is independent of the task object shape to both the task object and the destination, indicating the desired pose. This was inspired by the *Handles* visualization of Sukan et al. [41] and the *Axes* visualization of Andersen and Popescu [1], both of which use visualizations independent of the shape of the task object. We found that the conventional *xyz*-axes gnomon is geometrically symmetric and even though the axes may be colored red, green, and blue (e.g., as in Unity), it can be visually ambiguous, slowing down the user when they are trying to perform quickly. We saw this in our internal testing.

An example of a goal-based visualization is shown in Figure 3(a–c). One "Z" shape is placed on the task object and another on the destination. The user must align the two by translating and rotating the task object. Each "Z" is made from two differently colored "L"s of different sizes to aid users in distinguishing between the two, even if they have a color-vision deficiency. A small green sphere is placed where the two "L"s connect to help the user recognize the center of the visualization.

We use bright cyan and bright red for the "Z" on the current task object, and dark cyan and dark red for the "Z" at its destination. Note that as long as the orientation difference between the two visualizations is equal to the required rotation, the user will be able to finish the rotation with the visualization. This means we can rotate the visualizations in multiple ways. We created three


Figure 3: Goal-based visualizations. (a–c) Goal-based-Independent. The visualization on the task object represents its orientation, and the one on the destination shows the 6DoF goal pose. (a) The starting state of a step. (b) The user translates and rotates the object, with the visualization on the object updating accordingly while the one on the destination remains fixed. (c) The user aligns both visualizations to complete the task step. (d) Goal-based-Euler. (e) Goal-and-Action-based.

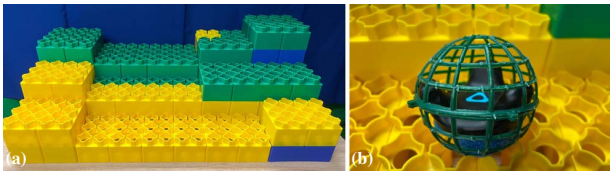


Figure 4: (a) Our AR testbed. (b) A task object.

versions of goal-based visualization that differ in how we placed the visualizations: Goal-based-Independent, Goal-based-Euler, and Goal-and-Action-based.

**Goal-based-Independent (*GI*)**: In GI (Figure 3a–c), the red "L" represents the top direction of the task object, and the cyan "L" represents the bottom direction. The orientation of the "Z" at the destination is computed based on the orientation of the "Z" on the task object and the required rotation. Note that the orientation of the "Z" on the task object is independent of the required rotation. We created GI to determine whether users can perform better than AE when given a suboptimal orientation.

**Goal-based-Euler (*GE*)**: In GE (Figure 3d), an orientation offset is added to the two "Z"s to align them with the Euler axes, computed in the same way as AE. We designed it to see whether providing a better orientation visualization for rotation improves performance. The users can compute the required rotation based on the orientations of the shorter bars perpendicular to the axes. We also compare whether the implicit indication of the required rotation provided by the Euler axis is more effective than an explicit circular arrow.

**Goal-and-Action-based (*GA*)**: GA (Figure 3e) is similar to GE, but includes the same circular arrow used in AE. We designed it to test whether the circular arrow improves performance compared to GE or if it confuses users.

Note that GI is purely goal-based, while GE and GA contain some action-based information: GE implicitly hints at the action, while GA includes the circular arrow used in AE to show the action explicitly. However, we still regard them as goal-based because we expect that the goal-based "Z"s will dominate the user performance.

For the action-based and goal-based precue visualizations, the colors brown and purple are used instead of red and cyan to avoid confusion with the cue visualizations.

### 4.3 Line Visualizations for Guiding Translation

Previous work has shown that line visualizations can communicate translation information effectively [20,48]. Therefore, after verifying this for our task through internal testing, we decided to use line visualizations similar to those of Liu et al. [20] to supplement the action-based and goal-based visualizations. We use a cyan line for each of the first two steps whose translation information is shown (Figure 1), connecting an object to its destination. The cyan line for the upcoming step is dimmer and thinner (1.0cm vs. 1.2cm) than the one for the current step. In addition, we add a yellow line that connects a wrist-worn tracker to the current task object to help the user find it (Figure 1a). The yellow line is hidden when the wrist-worn tracker is less than 20cm from the center of the current task object (Figure 1b). We did not add a yellow line connecting the first destination and the second origin, as internal testing did not find it beneficial.

### 4.4 Test Environment

We constructed a test environment using toy building blocks with dimensions of 10.16cm×10.16cm×6.35cm and 10.16cm×20.32cm×6.35cm. The final testbed is 101.6cm×50.8cm, with portions at five different heights (Figure 4a). We partitioned the test environment into 15 sections, each measuring either 20.32cm×15.24cm or 20.32cm×20.32cm. The center of each section serves as a potential target parallel to the $xz$-plane, which can act as either the location of an object or its destination. The test environment is mounted on a 121.92cm×60.96cm rectangular table, 45.72cm high, enabling easy access to any target.

Five task objects were each constructed from a Vive Tracker 2.0 placed in a 10cm-diameter plastic spherical cage, padded underneath by foam to center the tracker in the cage (Figure 4b). This makes it possible for the task object to be stably placed on the testbed at any orientation (which could not be done with a Vive Tracker 2.0 by itself, because of its irregular shape). Our user interface and associated application interface were developed using Unity 2019.4.18f1 [45] and the Mercury Messaging framework [7].

## 5 User Study

We conducted a user study to compare our visualizations and determine the number of precues that might benefit users. We also conducted two initial pilot studies before the user study to refine the visualizations and determine the appropriate length of our user study. The review board at our institution approved our studies.

### 5.1 Initial Pilot Studies

In a first pilot study with four participants, we tested early versions of the action-based and goal-based visualizations. We discovered that the circular arrow should avoid the $-y$ direction of the task object, as a user's hand might collide with the testbed if they try to follow it. Additionally, we found that using colors with the same hue but different brightness for the rotation cue and precue was confusing, and participants sometimes moved the current task object to the wrong destination. Consequently, we decided to use colors with different hues for the action-based and goal-based cue and precue visualizations.

In a second pilot study with three participants, we aimed to determine an appropriate length for the formal study. We included 27 trials, but participants were unable to finish in time and expressed frustration. As a result, we shortened the study length. We also discovered that participants performed much more slowly in the first two trials due to the task's difficulty and unfamiliarity. Thus, we decided to add practice trials to the formal study.

### 5.2 Hypotheses

After analyzing the results of our initial pilot studies and considering our initial design goals, we formulated the following hypotheses:

**H1.** *Participants will perform faster with GI, GE, and GA than with AE.* Goal-based visualizations make the end state clear. The user knows what they are aiming for and when they are done. Goal-based visualizations also allow the user to choose the action that is comfortable for them. This is especially important in our 6DoF interactions because the user's wrist must twist. In contrast, the suggested action in the action-based visualization might interfere with the user's mental planning.

**H2.** *Among the goal-based visualizations, the ones with "Z"s aligned with the Euler axis (GE and GA) will help participants perform faster than the one in which the "Z"s were not aligned with the Euler axis (GI).* In GE and GA, the user does not need to compute the rotation to align the long bars in the "Z"s as they are already initially parallel. The extra short bars perpendicular to the long bars, which suggest physical door-handle levers, make it easy to compute the rotation mentally.

**H3.** *Participants will not perform faster with the GA visualization compared to the GE visualization.* We believe that aligning the long bars of the "Z"s with the Euler axis will be sufficient to show the rotation information, while the added circular arrow of GA could distract the user, since it increases the number of components the user needs to track mentally.

**H4.** *Participants will perform faster with a precue compared to no precue.* Previous work [19–21, 48] showed that using a precue

could cause users to perform faster than with no precue. We believe that in our task the precue will also provide additional information to participants and help them prepare for the next step, resulting in a faster completion time.

## 5.3 Methods

### 5.3.1 Participants

We recruited 20 right-handed participants (8 female) aged 18 to 33 years (average 23.4). Participants were recruited via emails sent through our department email lists and posted flyers. Four of the participants had no prior experience with AR/VR, five had used AR/VR several times, four had used VR in class projects, three owned VR headsets for gaming, and four used VR for long-term jobs or research. Each participant received a 15 USD gift card as compensation. No participant in our formal study took part in the initial pilot studies.

### 5.3.2 Equipment

Study participants wore a Varjo XR-3 video–see-through AR headset [47] with a 115° horizontal field of view (134° diagonal at 12mm eye relief) and 90Hz refresh rate. The XR-3 was run in its outside-in tracking mode, tracked by four HTC SteamVR Base Station 2.0 units. Each participant wore a Vive Tracker 3.0 on their dominant hand, the only hand they were allowed to use for the study task. Five Vive Tracker 2.0 units were used as tracked objects, tracked using the same base stations as the XR-3. The XR-3 was driven by a computer with an Intel® Core™ i9-11900K processor and Nvidia GeForce RTX 3090 graphics card. Before each session, the headset, trackers, and table were sanitized using 70% isopropanol and the headset was also sanitized using a Cleanbox CX1 [6] UVC system.

### 5.3.3 Study Design

The study had eight conditions: (AE, GI, GE, GA) $\times$ (0, 1 Precue). We tested both visualization style and number of precues in the same study since previous research indicated there might be interactions between them [19]. The study aim was to find the best visualization style, so we blocked conditions by number of precues rather than visualization style, counterbalancing block order and visualization style order within blocks. We did this to avoid any participant encountering a specific style only at the beginning, middle, or end of the study.

Each block consisted of two timed trials, each consisting of a 16-step sequence (task). The order in which the blocks appeared was counterbalanced based on the condition to which each block belonged, ensuring each participant experienced a different order of conditions. Since we found in our initial pilot studies (Section 5.1) that participants were much slower in the first two trials, we added three practice trials at the beginning of the study.

Each of the 16 steps included the four phases described in Section 3. The users were required to complete each step before moving on to the next one. No object was cued and then precued in two subsequent steps. Each object was used in three to five of the 16 steps. The sequences were designed to form a closed loop (in terms of object position and orientation). This means we were able to pick a random start point in the sequence while ensuring that when a specific step was performed, the user would see the same upcoming translations and rotations.

To ensure the correct positions and orientations of the five objects at the beginning of each trial, we incorporated five setup steps where participants were required to move the objects to their designated starting poses. The visualizations functioned in the same manner during the setup steps as they did in the subsequent 16 regular steps, but the setup steps were excluded from the analysis.

To eliminate any potential confounding effects arising from participants anticipating the completion of a trial, we introduced a set of additional steps beyond the last step that needed to be completed. This approach allowed our precueing system to generate extra visualizations that extend beyond the actual end of the timed trial. Consequently, unless participants were consciously keeping track of the steps, which was not observed in initial piloting, they would not anticipate the trial's conclusion, since each step in a trial had the same number of precues.

### 5.3.4 Procedure

Each participant was first welcomed by the study coordinator and presented with an information sheet. After giving informed consent, the participant was then introduced to the study flow and given the Stereo Optical Co. Inc. Stereo Fly Test (SFT) [38], which contains nine questions, to screen for stereo vision, the Ishihara Pseudo-Isochromatic Plate (PIP) test [16] to screen for color deficiencies, and the Vandenberg–Kuse Mental Rotation Test (MRT) [46], which contains 20 questions, to screen for spatial ability. One participant answered four of the SFT questions correctly, two participants answered six correctly, three answered eight correctly, and the rest answered all correctly. For the PIP test, one participant had a color-vision deficiency and could not answer any question in the PIP test correctly, and the rest answered all correctly. In the MRT, the median score was 10.5, while the highest and lowest scores were 19 and 5, respectively. While the SFT, PIP, and MRT results were not used to determine eligibility for the study, we use them in Section 6.1 to help explain the performance differences between participants.

The study coordinator next put the headset on the participant and attached a Vive Tracker 3.0 to the participant's dominant hand using a Rebuff Reality TrackStrap. The study coordinator then started the study program. After starting the program, the study coordinator first calibrated the target positions to the height of the table. Then the participant entered a "sandbox mode," in which the study coordinator explained the study mechanism and how to follow the visualizations. After finishing the sandbox trials, the participant proceeded to the timed trial blocks.

Throughout the study, the headset and trackers tracking data and the trial and step completion times were recorded. Over the session, the participant's interaction was monitored by the study coordinator through a separate display.

After finishing all trials, the participant was asked to fill out a questionnaire that included questions on their demographics, an unweighted NASA Task Load Index (TLX) survey [13], and a request to rank the techniques based on their effectiveness. The TLX was modified to use a 1–7 scale [23], with 1 as best, rather than the original 0–20 scale. Each participant rated the visualizations for each TLX metric. We did not distinguish GI and GE in the questionnaire as they were quite similar. Images of each visualization were displayed during the rating process to remind participants of the visualizations used. Giving only a single questionnaire at the end of the study helped reduce the length of the study, and avoided the issue that the participant's criteria for answering the questionnaire might shift over time. The whole process took about 70 to 90 minutes for a typical participant to complete.

# 6 Results

## 6.1 Task Completion Time

The purpose of our study was to investigate the effectiveness of each visualization and the impact of number of precues, as measured by step completion time. As indicated in Section 3, we verified step completion by checking whether the task object had a pose within 3cm and 15° of the designated pose, using the same positional and rotational tolerances as Liu et al. [20]. To calculate step completion time, we recorded the time from when the 6DoF pose of the previous step was matched to when the 6DoF pose of the current step was matched. In our calculations, we included the time taken for a user to deposit an object in the subsequent step rather than the current step. We made this decision because it was difficult to determine

Table 1: Formal study step-completion time.

| | AE | GI | GE | GA |
|---|---|---|---|---|
| 0 Precues | 4.674 | 4.023 | 3.784 | 3.900 |
| 1 Precue | 4.876 | 4.162 | 3.918 | 4.008 |

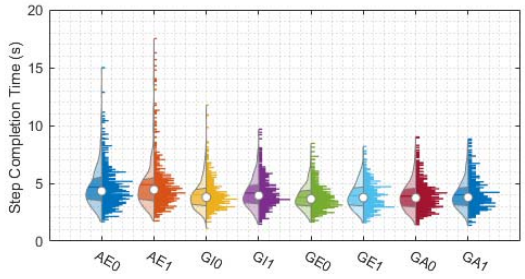


Figure 5: Distribution of step-completion time in each condition. The number in the condition name indicates the number of precues.

solely based on hand-position information whether the user had successfully deposited the object. Depositing the object might have been accomplished using only the fingers rather than the entire hand. For a detailed analysis of the various types of errors and their frequencies in each condition, please refer to Section 6.3.

Once the study was concluded, we processed the completion-time results generated automatically by our system before analyzing them. To identify outliers, we applied Tukey's outlier filter [44], computing the *outside fence* individually for each condition and user. Steps that exceeded the value of the third quartile plus 1.5 times the interquartile range were considered outliers. We anticipated that the different conditions would have a significant impact on completion time, and we observed variations in performance among different users. For the percentage of steps labeled as outliers in each condition, please refer to the supplementary material.

The average step completion time after outlier removal under each condition is shown in Table 1 and the distribution can be found in Figure 5. Note that the step completion time does not follow a normal distribution, as it was affected by the step difficulty and individual user performance, and error bars were not meant to show significance in differences between conditions. To better show between-subject differences, we plotted individual participant performance by visualization type and by number of precues. The plot can be found in the supplementary materials.

All hypotheses were evaluated with a significance level of $\alpha = .05$. We employed the MATLAB Statistics and Machine Learning Toolbox [22] to fit a linear mixed-effects model to our dataset. The step completion time served as the observation, while the fixed-effect variables included the visualization type and the number of precues. Additionally, the random-effect variables consisted of step (defined as the required 3D translation and 3D rotation in that step) and user ID. There is no interaction term between random-effect variables. We chose these variables after comparing the current model to alternative ones using a likelihood ratio test, which also involved Akaike and Bayesian information criteria. Specifically, we compared two models, one with interaction terms between the fixed-effect variables and one without. The results of the likelihood ratio test showed that both models fit the data equally well, indicating a weak interaction between the fixed-effect variables. As such, we selected the model without interaction terms between the fixed-effect variables. By comparing different linear mixed-effects models, we found that neither AR/VR experience, nor color-vision deficiency, nor stereo vision are significant factors, while MRT score is. Participants with higher MRT scores performed faster than those with lower scores. However, using user ID instead of MRT score as the random-effect term made the model fit the data better. This may be because the user ID random-effect term could absorb the impacts of MRT score, AR/VR experience, color-vision deficiency, stereo vision, and even factors affecting user performance but not measured in our study.

The $p$-values for the fixed-effect terms are $< .0001$, providing strong evidence for the significant effects of visualization type and number of precues. The model summary can be found in the supplementary material. The model also yielded large effect sizes, with $\eta^2 = 0.5598$ and Cohen's $d = 0.9695$ [5].

Overall, our analyses support **H1**, **H2**, and **H3**, but not **H4**.

To test **H1**, we analyzed the $p$-values of the fixed-effect terms for visualization type in our linear mixed-effects model. The $p$-values for the goal-based visualizations (GI, GE, and GA) were all $< .0001$, indicating that participants completed the steps faster with them compared to AE. Therefore, our results support **H1**.

To test **H2**, we compared performance using GI to performance using GE and GA. We calculated the contrast between GI and GE, and the contrast between GI and GA. The resulting $p$-values were both $< .0001$, indicating that GA and GE led to significantly better performance than GI. Thus, **H2** is supported.

To evaluate **H3**, we compared the performance of participants using GA and GE. We calculated the contrast between the two and found a $p$-value of .0239. This indicates that adding the arrow in GA significantly reduced performance when compared to GE. Therefore, **H3** is supported.

To evaluate **H4**, we analyzed the completion time data presented in Table 1. Contrary to our hypothesis, we found that participants were significantly slower with 1 Precue than with 0 Precues. Thus, **H4** is not supported. The $p$-value for the precue term is $< .0001$, indicating that the precue had a negative impact on participant performance.

Some participants achieved better results with a precue, as can be seen in the individual participant performance plot in the supplementary materials. To assess the statistical significance of this observation, we employed a linear mixed-effects model. The model incorporated User ID and number of precues as the fixed-effect variables, and included an interaction between these variables. Additionally, we used step as the random-effect variable. The results indicate that using a precue significantly reduced the step completion time for two participants, P6 and P10, with $p$-values of 0.0071 and 0.0003, respectively. These findings suggest that a precue benefited only a few participants. Moreover, the model revealed large effect sizes, with $\eta^2 = 0.6000$ and Cohen's $d = 1.0392$ [5].

To avoid type-I errors, we used the Holm–Bonferroni method [15]. We checked a total of six $p$-values for the validated hypotheses (three for **H1**, two for **H2**, and one for **H3**). Five of them are $< .0001$, and the remaining one is .0239. With this order, the $p$-values are smaller than $.05/6$, $.05/5$, ..., $.05/1$, respectively, meaning they survive their corresponding Holm–Bonferroni-corrected $\alpha$.

### 6.2 User Feedback

The unweighted NASA TLX results for our study are displayed in Figure 6, along with the $p$-values for each metric calculated using Friedman tests. The results indicate that the differences in mental demand, physical demand, performance, effort, and frustration are statistically significant. However, there is no significant difference in temporal demand ratings between the visualizations. It is worth noting that although the participants' rating on temporal demand is not significant, the objective results in Table 1 show that they performed faster with GI/GE than GA, as noted in Section 6.1.

The participants were asked to rank different visualization styles based on their preferences (Figure 7). The results show that GA was the most preferred style, followed by GI/GE, and finally AE. We ran a Friedman test on the preference data, which shows that

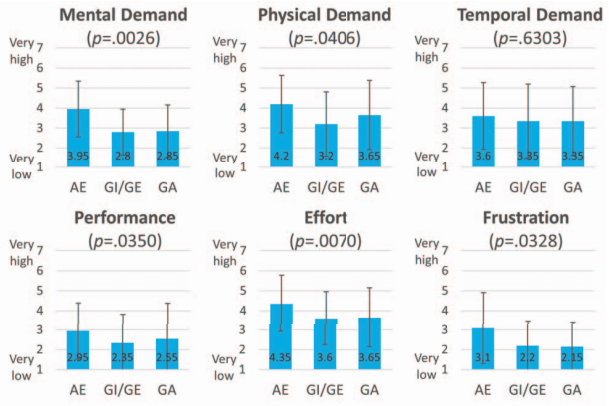


Figure 6: Unweighted NASA TLX results.

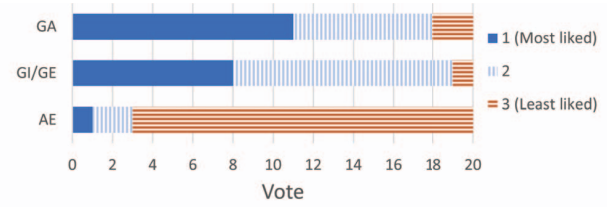


Figure 7: Preferences for visualization styles.

the differences in preference are significant ($p < .0001$). We further analyzed the data using Wilcoxon signed-rank tests to evaluate the difference in user preference between pairs. The results indicate that the user preference between GI/GE and GA was not significant with a $p$-value of .6972. However, the user preference for GI/GE and GA was significantly higher than for AE, with $p$-values of .0004 and .0011, respectively.

We also asked participants whether the precue was useful. Seven participants reported that they could use only one precue, while the remaining 13 participants reported that they could use only the cue.

### 6.3 Error Analysis

We analyzed the steps identified as outliers using Tukey's method [44] to gain insights into the errors made by participants in the study. Remember that even if a participant made an error, they were required to complete the current step before moving on to the next step. Given that there could be various underlying factors contributing to these errors, we manually labeled them for further analysis. We identified the following major types of errors made by participants and used only the closest label for each step with multiple errors:

**Inefficient Grasp (1.48%)** In these cases, the participant had difficulty completing the rotation due to a faulty grasp. They were unable to continue rotating the task item without first releasing it and re-grasping it, as evidenced by inconsistent hand and tracker movement traces.

**Tracking (0.47%)** On rare occasions, tracking failure occurred, which resulted in users spending more time on certain steps.

**Wrong Object (0.47%)** In some cases, the participant's hand moved to an incorrect task object, one that was not cued or precued.

**Work on Precue (0.33%)** The participant moved their hand to a precued task object instead of the cued one.

**Dropped Object (0.14%)** The participant accidentally dropped the task object, causing it to fall out of the testbed. This was confirmed by detecting that the tracker had moved outside of the designated testbed volume.

**Others (1.31%)** This category includes errors that do not fit into the previously defined types. Some instances involve participants spending a significant amount of time planning a step, despite no abnormalities observed in the hand-motion path or tracker rotations.

The supplementary material provides the percentage of each error type under each condition. Chi-square tests were conducted to determine if participants made more errors of a specific type under different conditions. The results indicate that all comparisons yielded either an insignificant $p$-value ($> .05$) or a small effect size ($\phi < 0.1$). This suggests that participants' completion time was influenced more by the experimental conditions than error rate.

## 7 Discussion

### 7.1 Comparison between Visualization Styles

The results of our study support **H1**. Participants achieved better performance in our 6DoF task with visualizations depicting the goal rather than the actions. Previous studies on cueing and precueing for translation tasks [19, 48] have shown that providing a path to the destination using a line (which is similar to providing the action) is more helpful than providing information about the destination using a number or a circle (which is similar to providing the goal). One possible explanation for this difference is that translation tasks are relatively easy and people are familiar with following lines. In contrast, 3D rotations are difficult and hard to imagine mentally [17, 31]. In addition, visualizations of 3D rotations are difficult to understand and instantiate. For 3D rotations, it seems to be easier for people to figure out the actions needed to achieve the desired end state than to understand and enact visualizations of 3D rotations. Interestingly, research investigating imitation in young children has found that they frequently imitate the goals but not necessarily the specific actions they saw reach the goal [9, 11, 24, 27].

Moreover, our study provides evidence for **H2** that configuring the goal-based visualizations to align with the Euler axis enhances user speed. This may be because parallel alignment between the visualizations on the task object and the destination enables users to compute the required rotation based on the orientations of the shorter bars perpendicular to the axes in GE and GA, while also hinting at the proper grasping pose to hold the sphere.

Furthermore, the performance comparison between GE and AE in Table 1 suggests that letting users compute the required rotations by aligning the shorter bars perpendicular to the axes is more effective than providing circular arrows indicating suggested rotations. This may be because merging the axes and shorter bars in GE allows users to track them as a whole rather than as separately as in AE.

In regard to **H3**, the comparison between GE and GA reveals that participants performed more slowly in GA, indicating that the additional arrows did not benefit users. While some participants preferred GA according to the feedback in Section 6.2, the completion time data did not support this preference. Therefore, using the GE visualization should be more efficient.

### 7.2 Number of Usable Precues

Previous studies have shown that precues can be effective in enhancing performance in several tasks requiring a sequence of moves [19–21, 48]. Liu et al. [19] demonstrated that participants were able to use two to three precues with effective precueing styles in a translation-only task. However, in tasks that involve translation and rotation, Liu et al. [20] found that participants were not able to use rotation precues, but could use translation precues, while Liu et al. [21] demonstrated that a single rotation precue could benefit users in rotation-only tasks. In contrast, our evaluation of **H4** suggests that most participants did not benefit from even a single precue in our task, which required both translation and rotation, as discussed

in Section 6.1. This may have been because the task was challenging, and the precue might have overloaded information processing capacity in many participants, negating any potential benefits. The greater difficulty of our task than the one in Liu et al. [21] is evident in the step completion time. We adopted the same target tolerances of 3cm and 15°, but the average step completion time was less than 2s in Liu et al. [21], while ours has an average step completion time of more than 3.7s. Nevertheless, P6 and P10 in the formal study (and the first author when tested separately) performed faster with a precue, indicating that individual abilities and training might play a role in determining the usefulness of a precue.

Another difference between our study and the previous work is the size of the testbed. The testbeds used in the previous studies [19–21, 48] were smaller than 35cm×35cm, which made it easier for participants to track all the visualizations in their near-peripheral vision, assuming an average arm length of 60cm and a near-peripheral vision range of 30° [40]. In contrast, the present testbed was larger, measuring approximately 100cm×50cm, which made it more likely for the task object to fall out of the participants' near-peripheral vision. To compare our results with the ones of Liu et al. [19, 20], we removed the steps in which the origins and destinations of the current and the first upcoming steps did not fall in a 35cm×35cm window, and reran the linear mixed-effects model in Section 6.1. The result still shows that adding a precue increased completion time ($p = .001$). This argues against the increase in completion time being caused by the greater distance between the cue and the precue in our study.

### 7.3 Follow-Up Pilot Study on the Efficacy of Precueing

To further investigate whether modifying our precue visualizations could make them more useful, we conducted a small-scale follow-up pilot study with four participants, approved by the review board at our institution. Two participants tested whether including a line connecting the current destination to the origin of the next step would enhance user performance. The other two participants tested whether simplifying the precue by showing only part of it, such as just the line or the origin copy of the visualization but not the destination copy, would improve performance. The intent was to reduce the amount of information users must process and keep in mind. However, the results did not support either variant improving performance. This suggests that participants' inability to utilize the precue may be attributed to the difficulty of the task and the added burden of processing the precue.

### 7.4 Lessons Learned

From our user study, we found that displaying an unambiguous shape indicating the goal of the step improved user performance. Moreover, suggesting a rotation based on Euler's rotation theorem in the goal-based visualization further enhanced performance. Based on these findings, we suggest that when creating visualizations for cueing sequential 6DoF rigid-body transformations, designers should consider using distinct unambiguous shapes to display rotation information, including each task object's orientation goal, while leaving the user to compute the action to achieve the goal.

In addition, the effectiveness of precues appears to be influenced by task difficulty and individual ability. Our task uses spherical task objects and any possible rotation direction, and our visualizations are designed to work independently of the shape of task objects. However, in real-life scenarios, the shapes of objects may restrict the range of rotation, potentially simplifying the task. Moreover, trained users may benefit from precues. Thus, information about the specific task and users may determine whether precues would be effective.

## 8 Limitations and Future Work

### 8.1 Participant Population and Laboratory Setting

The participants in our study were recruited from our institution, and all were relatively young. However, it is important to note that further research is required to establish whether our findings are applicable to older people. Moreover, though we did not design our study targeting specific handedness, our participants were all right-handed. Future work could address whether handedness might have an impact on performance. Furthermore, our study employed a single-session design, and thus did not measure the performance of trained users. This may have contributed to the observed disparity between the first author and most participants.

It is also worth noting that our work was done in a controlled laboratory setting. In real-world applications, users might perform differently. We would like to address these limitations in the future.

### 8.2 Cueing Grasping Pose Based on Ergonomics

Our current AE, GE, and GA visualizations implicitly cue grasp poses based on the rotation computed with Euler's rotation theorem. The user can put their hand in a convenient pose to rotate the task object around the Euler axis. However, this approach does not consider factors such as the object's pose relative to the user, the user's handedness, and a human arm model. Although previous work has explored grasping items [3, 34, 39, 51], it would be interesting to investigate how these factors could be incorporated into the cues and precues. Another possible direction is to show different cues based on how a user interacts with the task item, similar to the work by Satriadi et al. [33]. Moreover, our study focuses on a unimanual task. Future work could explore cueing for sequential 6DoF bimanual tasks.

## 9 Conclusions

Our study focused on assisting users in performing 6DoF rigid-body transformations using graphical cues in a sequential task. In each step, the user had to pick up a physical object, rotate it in 3D while translating it in 3D, and deposit it in a target 6DoF pose. We designed an action-based visualization that explicitly shows suggested actions for users to complete the task, as well as goal-based visualizations that explicitly communicate the goal of each step. We compared the effectiveness of the action-based visualization and three variations of the goal-based visualization in a study that evaluated the number of precues users could use. Our results indicated that the goal-based visualizations, particularly those with "Z"s aligned with the Euler axis computed based on the task object orientation and goal pose, resulted in faster task completion than the action-based visualization. However, adding circular arrows used in the action-based visualization to the goal-based visualizations did not further improve performance.

Interestingly, our study showed that precues did not provide a benefit to most of the participants, possibly due to the difficulty of the task. Many participants found the precues distracting, which may be attributed to the single-session design of the study. However, two of the participants and the first author were able to benefit from a precue. Nonetheless, our findings are relevant to a variety of real-world tasks, such as maintenance and repair, that involve sequential 6DoF rigid-body transformations.

## Acknowledgments

This research was funded in part by National Science Foundation Grant CMMI-2037101. We thank Shutaro Aoyama for his assistance in making the video and Long Zhao, Joe Suk, and Harry Xi for their feedback on our statistical analysis.

## REFERENCES


[1] D. Andersen and V. Popescu. AR interfaces for mid-air 6-DoF alignment: Ergonomics-aware design and evaluation. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 289–300, 2020. doi: 10.1109/ISMAR50242.2020.00055

[2] D. Andersen, P. Villano, and V. Popescu. AR HMD guidance for controlled hand-held 3D acquisition. *IEEE Transactions on Visualization and Computer Graphics*, 25(11):3073–3082, 2019. doi: 10.1109/TVCG.2019.2932172

[3] S. C. Bae. *Investigation of hand posture during reach and grasp for ergonomic applications*. PhD thesis, University of Michigan, 2011.

[4] H. Bai, P. Sasikumar, J. Yang, and M. Billinghurst. A user study on mixed reality remote collaboration with eye gaze and hand gesture sharing. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*, CHI '20, p. 1–13. Association for Computing Machinery, New York, NY, USA, 2020. doi: 10.1145/3313831.3376550

[5] M. Brysbaert and M. Stevens. Power analysis and effect size in mixed effects models: A tutorial. *Journal of cognition*, 1(1), 2018. doi: 10.5334/joc.10

[6] Cleanbox Technology. *CleanboxCX1*. `https://cleanboxtech.com/products/`, last accessed on 06/04/2023.

[7] C. Elvezio, M. Sukan, and S. Feiner. Mercury: A messaging framework for modular UI components. In *Proceedings of the 2018 CHI Conference on Human Factors in Computing Systems*, CHI '18, pp. 588:1–588:12. ACM, New York, NY, USA, 2018. doi: 10.1145/3173574.3174162

[8] L. Euler. Formulae generales pro translatione quacunque corporum rigidorum. *Novi Commentarii academiae scientiarum Petropolitanae*, 20:189–207, 1776.

[9] M. Gattis, H. Bekkering, and A. Wohlschläger. Goal-directed imitation. In A. N. Meltzoff and W. Prinz, eds., *The imitative mind: Development, evolution, and brain bases*, vol. 6, pp. 183–205. Cambridge University Press, 2002. doi: 10.1017/CBO9780511489969.011

[10] F. B. Gilbreth and L. M. Gilbreth. *Applied motion study: A collection of papers on the efficient method to industrial preparedness*. Macmillan, 1919.

[11] B. Gleissner, H. Bekkering, and A. N. Meltzoff. Children's coding of human action: cognitive factors influencing imitation in 3-year-old. *Developmental Science*, 3(4):405–414, 2000. doi: 10.1111/1467-7687.00135

[12] Y. Harazono, H. Ishii, H. Shimoda, Y. Taruta, and Y. Kouda. Development of ar-based scanning support system for 3d model reconstruction of work sites. *Journal of Nuclear Science and Technology*, 59(7):934–948, 2022. doi: 10.1080/00223131.2021.2018369

[13] S. G. Hart and L. E. Staveland. Development of nasa-tlx (task load index): Results of empirical and theoretical research. In P. A. Hancock and N. Meshkati, eds., *Human Mental Workload*, vol. 52 of *Advances in Psychology*, pp. 139–183. North-Holland, 1988. doi: 10.1016/S0166-4115(08)62386-9

[14] A. D. Hartl, C. Arth, J. Grubert, and D. Schmalstieg. Efficient verification of holograms using mobile augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 22(7):1843–1851, 2016. doi: 10.1109/TVCG.2015.2498612

[15] S. Holm. A simple sequentially rejective multiple test procedure. *Scandinavian Journal of Statistics*, 6(2):65–70, 1979.

[16] S. Ishihara. *Ishihara's Tests for Colour-blindness*. Kanehara Shuppan, 1972.

[17] R. L. Klatzky, J. M. Loomis, A. C. Beall, S. S. Chance, and R. G. Golledge. Spatial updating of self-position and orientation during real, imagined, and virtual locomotion. *Psychological Science*, 9(4):293–298, 1998. doi: 10.1111/1467-9280.00058

[18] C. Liu, S. Huot, J. Diehl, W. Mackay, and M. Beaudouin-Lafon. Evaluating the benefits of real-time feedback in mobile augmented reality with hand-held devices. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '12, p. 2973–2976. Association for Computing Machinery, New York, NY, USA, 2012. doi: 10.1145/2207676.2208706

[19] J.-S. Liu, C. Elvezio, B. Tversky, and S. Feiner. Using multi-level precueing to improve performance in path-following tasks in virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 27(11):4311–4320, 2021. doi: 10.1109/TVCG.2021.3106476

[20] J.-S. Liu, B. Tversky, and S. Feiner. Precueing object placement and orientation for manual tasks in augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 28(11):3799–3809, 2022. doi: 10.1109/TVCG.2022.3203111

[21] J.-S. Liu, B. Tversky, and S. Feiner. Precueing sequential rotation tasks in augmented reality. In *Proceedings of the 28th ACM Symposium on Virtual Reality Software and Technology*, VRST '22. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3562939.3565641

[22] The MathWorks, Inc. *Matlab Statistics and Machine Learning Toolbox*. `https://www.mathworks.com/help/stats/index.html`, last accessed on 06/04/2023.

[23] MeasuringU. *10 Things to Know about the NASA TLX*. `https://measuringu.com/nasa-tlx/`, last accessed on 06/04/2023.

[24] A. N. Meltzoff and M. K. Moore. Explaining facial imitation: a theoretical model. *Early Development and Parenting*, 6(3-4):179–192, 1997. doi: 10.1002/(SICI)1099-0917(199709/12)6:3/4<179::AID-EDP157>3.0.CO;2-R

[25] P. Mijkesenaar and P. Westendorp. *Open here: the art of instructional design*. Joost Elffers Books, 1999.

[26] O. Oda, C. Elvezio, M. Sukan, S. Feiner, and B. Tversky. Virtual replicas for remote assistance in virtual and augmented reality. In *Proceedings of the 28th Annual ACM Symposium on User Interface Software 38; Technology*, UIST '15, pp. 405–415. ACM, New York, NY, USA, 2015. doi: 10.1145/2807442.2807497

[27] H. Over and M. Carpenter. The social side of imitation. *Child Development Perspectives*, 7(1):6–11, 2013. doi: 10.1111/cdep.12006

[28] L. M. Parsons. Imagined spatial transformations of one's hands and feet. *Cognitive Psychology*, 19(2):178–241, 1987. doi: 10.1016/0010-0285(87)90011-9

[29] L. M. Parsons. Temporal and kinematic properties of motor behavior reflected in mentally simulated action. *Journal of Experimental Psychology: Human Perception and Performance*, 20(4):709, 1994. doi: 10.1037/0096-1523.20.4.709

[30] J. Polvi, T. Taketomi, A. Moteki, T. Yoshitake, T. Fukuoka, G. Yamamoto, C. Sandor, and H. Kato. Handheld guides in inspection tasks: Augmented reality versus picture. *IEEE Transactions on Visualization and Computer Graphics*, 24(7):2118–2128, 2018. doi: 10.1109/TVCG.2017.2709746

[31] J. J. Rieser. Access to knowledge of spatial structure at novel points of observation. *Journal of Experimental Psychology: Learning, Memory, and Cognition*, 15(6):1157–1165, 1989. doi: 10.1037/0278-7393.15.6.1157

[32] S. Sadasivan, J. S. Greenstein, A. K. Gramopadhye, and A. T. Duchowski. Use of eye movements as feedforward training for a synthetic aircraft inspection task. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '05, p. 141–149. Association for Computing Machinery, New York, NY, USA, 2005. doi: 10.1145/1054972.1054993

[33] K. A. Satriadi, J. Smiley, B. Ens, M. Cordeil, T. Czauderna, B. Lee, Y. Yang, T. Dwyer, and B. Jenny. Tangible globes for data visualisation in augmented reality. In *Proceedings of the 2022 CHI Conference on Human Factors in Computing Systems*, CHI '22. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3491102.3517715

[34] D. Schön, T. Kosch, F. Müller, M. Schmitz, S. Günther, L. Bommhardt, and M. Mühlhäuser. Tailor twist: Assessing rotational mid-air interactions for augmented reality. In *Proceedings of the 2023 CHI Conference on Human Factors in Computing Systems*, CHI '23. Association for Computing Machinery, New York, NY, USA, 2023. doi: 10.1145/3544548.3581461

[35] J. Shingu, E. Rieffel, D. Kimber, J. Vaughan, P. Qvarfordt, and K. Tuite. Camera pose navigation using augmented reality. In *2010 IEEE International Symposium on Mixed and Augmented Reality*, pp. 271–272, 2010. doi: 10.1109/ISMAR.2010.5643602

[36] R. Sodhi, H. Benko, and A. Wilson. Lightguide: Projected visualizations for hand movement guidance. In *Proceedings of the SIGCHI*

*Conference on Human Factors in Computing Systems*, CHI '12, p. 179–188. Association for Computing Machinery, New York, NY, USA, 2012. doi: 10.1145/2207676.2207702

[37] A. Stanescu, P. Mohr, D. Schmalstieg, and D. Kalkofen. Model-free authoring by demonstration of assembly instructions in augmented reality. *IEEE Transactions on Visualization and Computer Graphics*, 28(11):3821–3831, 2022. doi: 10.1109/TVCG.2022.3203104

[38] Stereo Optical Co. Inc. *Original Stereo Fly Stereotest*. `https://www.stereooptical.com/products/stereotests-color-tests/original-stereo-fly/`, last accessed on 06/04/2023.

[39] F. Stival, S. Michieletto, M. Cognolato, E. Pagello, H. Müller, and M. Atzori. A quantitative taxonomy of human hand grasps. *Journal of neuroengineering and rehabilitation*, 16:1–17, 2019. doi: 10.1186/s12984-019-0488-x

[40] H. Strasburger, I. Rentschler, and M. Jüttner. Peripheral vision and pattern recognition: A review. *Journal of Vision*, 11(5):13–13, 12 2011. doi: 10.1167/11.5.13

[41] M. Sukan, C. Elvezio, S. Feiner, and B. Tversky. Providing assistance for orienting 3D objects using monocular eyewear. In *Proceedings of the 2016 Symposium on Spatial User Interaction*, SUI '16, p. 89–98. Association for Computing Machinery, New York, NY, USA, 2016. doi: 10.1145/2983310.2985764

[42] M. Tucker and R. Ellis. On the relations between seen objects and components of potential actions. *Journal of Experimental Psychology: Human Perception and Performance*, 24(3):830, 1998. doi: 10.1037/0096-1523.24.3.830

[43] K. Tuite, N. Snavely, D.-Y. Hsiao, A. M. Smith, and Z. Popović. Reconstructing the world in 3D: Bringing games with a purpose outdoors. In *Proceedings of the Fifth International Conference on the Foundations of Digital Games*, FDG '10, p. 232–239. Association for Computing Machinery, New York, NY, USA, 2010. doi: 10.1145/1822348.1822379

[44] J. W. Tukey. *Exploratory data analysis*, vol. 2. Reading, Mass., 1977.

[45] Unity. *Unity*. `https://unity.com/`, last accessed on 06/04/2023.

[46] S. G. Vandenberg and A. R. Kuse. Mental rotations, a group test of three-dimensional spatial visualization. *Perceptual and Motor Skills*, 47(2):599–604, 1978. doi: 10.2466/pms.1978.47.2.599

[47] Varjo. *Varjo XR-3*. `https://varjo.com/products/xr-3/`, last accessed on 06/04/2023.

[48] B. Volmer, J. Baumeister, S. Von Itzstein, I. Bornkessel-Schlesewsky, M. Schlesewsky, M. Billinghurst, and B. H. Thomas. A comparison of predictive spatial augmented reality cues for procedural tasks. *IEEE Transactions on Visualization and Computer Graphics*, 24(11):2846–2856, 2018. doi: 10.1109/TVCG.2018.2868587

[49] B. Volmer, J. Baumeister, S. Von Itzstein, M. Schlesewsky, I. Bornkessel-Schlesewsky, and B. H. Thomas. Event related brain responses reveal the impact of spatial augmented reality predictive cues on mental effort. *IEEE Transactions on Visualization and Computer Graphics*, pp. 1–18, 2022. doi: 10.1109/TVCG.2022.3197810

[50] A. Wohlschläger and A. Wohlschläger. Mental and manual rotation. *Journal of Experimental Psychology: Human perception and performance*, 24(2):397, 1998. doi: 10.1037/0096-1523.24.2.397

[51] X. Zhang, K. Ikematsu, K. Kato, and Y. Sugiura. Evaluation of grasp posture detection method using corneal reflection images through a crowdsourced experiment. In *Companion Proceedings of the 2022 Conference on Interactive Surfaces and Spaces*, ISS '22, p. 9–13. Association for Computing Machinery, New York, NY, USA, 2022. doi: 10.1145/3532104.3571457

[52] Y. Zhao, J. Stefanucci, S. Creem-Regehr, and B. Bodenheimer. Evaluating augmented reality landmark cues and frame of reference displays with virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 29(5):2710–2720, 2023. doi: 10.1109/TVCG.2023.3247078

[53] B. Zhou and S. Güven. Fine-grained visual recognition in mobile augmented reality for technical support. *IEEE Transactions on Visualization and Computer Graphics*, 26(12):3514–3523, 2020. doi: 10.1109/TVCG.2020.3023635