Precueing Object Placement and Orientation for Manual Tasks in Augmented Reality

Jen-Shuo Liu, Barbara Tversky, and Steven Feiner

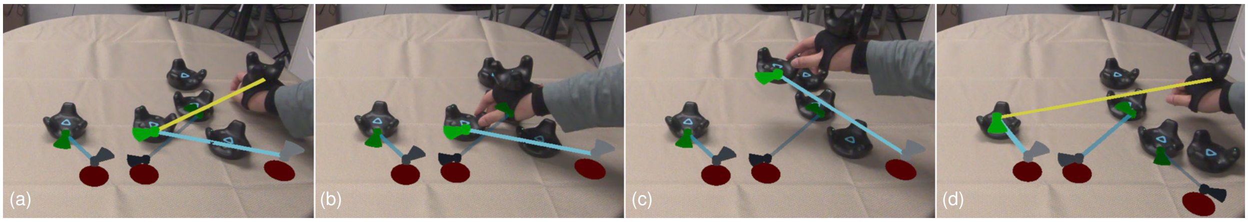

Fig. 1. A subtask in our study. The user is required to move their hand to pick up a designated object, transport and rotate it to a specified place and orientation, and deposit it. We add visualizations to cue and precue movement and rotation to guide the user. Cyan lines guide the user's hand to connect objects (Vive trackers) to their destinations (red circles) for current and future subtasks and wedge sets are used to cue and precue rotation. Each wedge set consists of a green wedge showing the current orientation of an object, and a gray one indicating the goal orientation. The four phases of a subtask: (a) The user's unloaded hand, guided by the yellow line, moves to the object to be manipulated. (b) The user is about to pick up the object and the yellow line has disappeared. (c) The user's loaded hand moves and rotates the object, transporting it to its destination. (d) The user deposits the object at its destination with the specified orientation. Their unloaded hand is now connected to the next object to be manipulated, and the visualizations are updated.

Abstract— When a user is performing a manual task, AR or VR can provide information about the current subtask (cueing) and upcoming subtasks (precueing) that makes them easier and faster to complete. Previous research on cueing and precueing in AR and VR has focused on path-following tasks requiring simple actions at each of a series of locations, such as pushing a button or just visiting. We consider a more complex task, whose subtasks involve moving to and picking up an item, moving that item to a designated place while rotating it to a specific angle, and depositing it. We conducted two user studies to examine how people accomplish this task while wearing an AR headset, guided by different visualizations that cue and precue movement and rotation. Participants performed best when given movement information for two successive subtasks and rotation information for a single subtask. In addition, participants performed best when the rotation visualization was split across the manipulated object and its destination.

Index Terms—Augmented reality, precueing, cueing, manual tasks

1 Introduction

Augmented reality (AR) and virtual reality (VR) have been used to guide users in a wide range of tasks, helping them perform more quickly and accurately. This can be done through *cueing*, that is, providing information (*cues*) about the subtask that is about to be performed. For example, Bai et al. [3] developed an AR system to help a remote expert show cues to a technician to help them pick up and place task objects, while Zhou and Güven [45] use computer vision to identify the phase of a task and automatically show cues for it to a technician.

Although cueing the current step of a series of actions is helpful, people can anticipate several steps in the future for practiced skills such as playing the violin, [19]. AR or VR systems can also guide the users by *precueing*, that is, providing information (*precues*) about future subtasks before the current subtask has been completed. For example, in an aircraft inspection training system developed by Sadasivan et al. [32], an expert's eye gaze was recorded and shown to a novice for training. Part of the eye-gaze trace was relevant to what the novice was currently doing, but the remainder of the visualized eye-gaze trace served as precues. Research has shown that precues can help improve performance in path-following tasks on desktops [11], in VR [20], and in AR [42], as described in Section 2. These papers all focused on relatively simple tasks: moving between stationary targets with a mouse, touchpad, hand, or hand-held controller, and, in some cases, pressing buttons on an input device or in the physical task domain.

Building on this work, we explore whether and how precueing in AR could benefit a more complex task. We present an AR testbed for a compound task that is richer than used in previous work. Our task consists of a series of subtasks. In each subtask, the user moves their empty hand to an object, picks it up, rotates and transports it to a destination, and deposits it there.

Using this testbed, we contribute the results of two formal studies. In the first study, we evaluate the effects on performance of different styles of visualization for rotation cues and precues when used with movement cues and precues. We also evaluate how performance is affected by the number of cue and precue visualizations for movement and rotation. In the second study, we further investigate the role of the best visualization used in the first study for rotation. Taken together, the results show that among the rotation cue and precue visualizations studied, one that splits information between the object being manipulated and the destination yields the shortest completion time. In addition, giving participants movement information for two subtasks and rotation information for a single subtask yields the shortest completion time.

2 Related Work

Cues are widely used in AR and VR. For example, to assist in navigation, Thomas et al. [38] and Narzt et al. [25] use cues to show the user

- *Jen-Shuo Liu is with Department of Computer Science, Columbia University. E-mail: jl5004@columbia.edu.*
- *Barbara Tversky is with Department of Human Development, Teachers College, Columbia University. E-mail: bt2158@tc.columbia.edu.*
- *Steven Feiner is with Department of Computer Science, Columbia University. E-mail: feiner@cs.columbia.edu.*

Manuscript received 11 March 2022; revised 11 June 2022; accepted 2 July 2022. Date of publication 01 September 2022; date of current version 03 October 2022. Digital Object Identifier no. 10.1109/TVCG.2022.3203111

Table 1. Comparison of previous work on cueing and precueing.

	[25, 38]	[17, 18] [15, 33]	[28]	[2, 12] [5, 16]	[11, 42] [20]
Cue body destination	✓				
Focus on hands		✓	✓		✓
Cue rotation			✓		
Use gaze				✓	
Collaborative tasks				✓	
Precue task info					✓

where they need to go. These projects cue a user to a destination but do not focus on their hands.

Other work uses AR cues to guide the user's hands. For example, Klinker et al. [17] use AR to cue workers about which components to work on in power-plant maintenance. Jeanne et al. [15] compared hand-guiding techniques in VR and found spheres to be superior to arrows or paths in visualizing hand-guidance errors. Kosmalla et al. [18] investigate different cues for rock climbers. Sodhi et al. [33] project cues directly on a user's hand. These projects guide hand movement but do not address rotation. In contrast, Oda et al. [28] direct the user's hands to perform actions that include rotation. Their system supports a remote expert interactively assisting a local technician in AR, cueing the current task for the technician to perform. They focus on cueing the current subtask rather than precueing future subtasks.

Eye gaze, like pointing, is a cue to focus attention and, in context, intention to act [29]. This has been used in many AR/VR systems. Andrist et al. [2] incorporate a virtual agent who specifies an object by looking at it using eye gaze or head orientation. Higuch et al. [12] use the gaze of a remote expert to guide a local worker. In the system developed by Jing et al. [16], eye gaze supports collaborating users in searching and puzzle-solving. When they use gaze to communicate with each other, their gaze serves as cues for the task. Chen et al. [5] compare several ways to visualize eye gaze with AR, finding that a visible eye-gaze ray increased task speed. These systems use eye gaze or head orientation in collaborative tasks, focusing on speed and communication between users, but do not cue rotation. In addition, they focus on cueing the current subtask rather than precueing future subtasks.

Some researchers provide precues as well as cues to guide a series of actions. Hertzum and Hornbæk [11] ask participants to alternately tap/click a constant (hence precued) center target and one of a circular set of surrounding targets. They find that participants move faster to the precued center than to the cued peripheral targets. Volmer et al. [42] also use a single precue. In their case, the target varies, so the precue is informative and speeds performance. More recently, Liu et al. [20] use multiple precues in a VR path-following task. They find the best performance for two to three precues that use lines to connect targets.

All of these studies of precueing [11, 20, 42] use relatively simple path-following tasks. In contrast, we address how precueing in AR can be used in a significantly more complex, heterogeneous physical task. Table 1 summarizes the work we have discussed in this section.

3 Task and Visualizations

3.1 Task

We designed a heterogeneous task [21] that is composed of a series of subtasks. In each subtask, the user needs to move their hand to a designated item, pick it up, move it to a specified destination while rotating it by a specified angle, and deposit it. Figure 1 shows an example of one of the subtasks from which our task is composed. Using the distinction between an unloaded (empty) and loaded (full) hand made by Gilbreth and Gilbreth [10], each subtask can be divided into the following four phases (Figure 1):

Unloaded hand: At the beginning of each subtask, the user needs to move their unloaded hand to the designated object (Phase 1: Figure 1a) and pick up the object. Their hand is now loaded (Phase 2: Figure 1b).

Loaded hand: The user must next move the object to the specified destination and accomplish the specified rotation (Phase 3: Figure 1c), and put down the object (Phase 4: Figure 1d). While movement and rotation of the object can be performed simultaneously, the current subtask must be completed before proceeding to the next subtask. Information about where to move and how much to rotate is shown using separate visualizations, as described in Sections 3.3 and 3.4. When the user places the object at the specified destination with the specified orientation, the subtask is complete and the unloaded-hand phase of the next subtask begins (Figure 1d).

3.2 Cueing and Precueing

Cues communicate information about the current subtask, while precues communicate information about future subtasks.

Cue: Cues guide the user to perform the current subtask. As the user finishes the kth subtask, the cue visualizations update to guide the user to perform the $k+1$st subtask.

Precues: For a condition with m precued subtasks, when the user is working on subtask k, the precued subtasks are subtasks $k+1$, $k+2$, $k+3$, ..., $k+m$. Once the user completes the current subtask k, the cued subtask becomes subtask $k+1$ and the precued subtasks become $k+2$, $k+3$, ..., $k+m+1$. This continues until the user finishes the entire task.

3.3 Line Visualizations as Movement Cue and Precues

Previous work by Volmer et al. [42] and Liu et al. [20] shows that visualizations that connect an origin to a destination with a line in path-following tasks can guide the users more effectively than visualizations that do not use a line. Therefore, after verifying this was also true for our study, we decided to use lines to show movement information (as opposed to rotation information). We use a cyan line for each subtask whose movement information is shown, connecting its object and destination. Based on pilot studies, we decided to use movement visualizations for up to four subtasks. The first line cues movement for the current subtask. The second to fourth lines precue future subtasks. Each line is successively dimmer. We also use width differences to help users differentiate the lines (from first to fourth: 1.2cm, 1cm, 0.8cm, and 0.6cm).

Unlike Liu et al., we do not put terminators (e.g., circles or arrowheads) at the ends of lines, as they might occlude our rotation cues and precues, described in Section 3.4. We put only red spots at the destinations where objects should go (below the cyan lines in Figure 1). In addition, we add a yellow line that connects the user's empty hand to the object to be picked up to help the user navigate to the first object. This cues the user to move to the first object but does not contain enough information to finish a subtask, as it does not lead to the destination. The yellow line is hidden when the hand is less than 15cm from the current object.

The user's hand needs to visit two places in each subtask: the object to be moved and its destination. This is different from the path-following task in the study by Volmer et al. [42] or Liu et al. [20]. In their studies, the user moves their hand to only one place in each subtask. Each line in our task cues or precues two places: the object to be moved and its destination. Therefore, when s subtasks are cued or precued with s cyan lines, the lines show information for $2s$ places. When the user's hand is unloaded, there will be one place cued and $2s-1$ *places* precued. When the user's hand is loaded, there will be one place cued and $2s-2$ *places* precued, since the first object is already in the user's hand. For example, in Figure 1, three cyan lines cue and precue the movement information of three subtasks and connect six places. For an unloaded hand (Figure 1a), the first object is the cued place, while the first destination, second object, second destination, third object, and third destination are the five precued places. For a loaded hand (Figure 1b), the first destination is the cued place, while the second object, second destination, third object, and third destination are the four precued places. We use this to explain our results and compare them with those of Liu et al. [20] in Section 8.1.

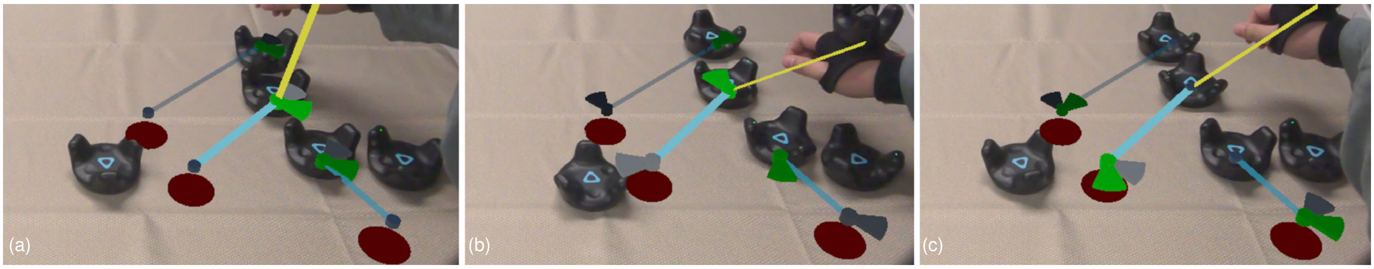

Fig. 2. Two-level precueing visualizations for an AR object movement and rotation task. Three wedge-set configurations tested in our user study: (a) both wedges on the object, (b) the green wedge on the object and the gray wedge on the destination, and (c) both wedges on the destination.

3.4 Wedge-Set Visualizations as Rotation Cue and Precues

Through pilot studies, we developed a *wedge-set* visualization to cue and precue object rotation. Each wedge set shows rotation information for one subtask: a green wedge shows the current orientation, while a gray wedge shows the target orientation.

Other visualizations tested in the pilot studies included an arrow–ring visualization [35] that we modified to use only one arrow ring to accommodate our use of 1DoF rotation. The arrow–ring visualization did not perform as well as the wedge-set visualization: when the user is working on the movement part of the subtask, it might be hard to check the orientation of the arrow. To address this, we use a larger simple shape instead of the ring. We also tested rubberband lines [28], using only a single rubberband line, which, together with the line visualization for movement, is sufficient to specify 1DoF rotation. This also performed worse than the wedge-set visualization: the rubberband line tended to confuse users when it intersected the movement cue. To address this issue, one can replace one of the lines with a shape such as a dot or a wedge.

Combining these solutions, we came up with the wedge-set visualization. Each wedge is a 45° circular sector, large enough to be easily observed, even when the user is doing a movement task. Its shape is different from the movement visualization, to minimize confusion. We used up to four wedge sets in our study. Like the lines showing movement, the wedge sets are successively dimmer. We also use size differences to help users differentiate the four wedge sets (the radii of the green and grey wedges are, from the first to fourth wedge set, 4.6cm, 4.1cm, 3.6cm, and 3.2cm).

One question that remains is where the two wedges should be placed. Figure 2 shows the three possible configurations we consider: (a) Object/Object (O/O), (b) Object/Destination (O/D), and (c) Destination/Destination (D/D). The O/O configuration is inspired by the arrow–ring visualization, where the visualizations are on the object itself. In contrast, the O/D configuration is inspired by the modified rubberband-line visualization. Finally, the D/D configuration addresses whether the user can focus on the destination and parse both movement information and rotation information at the same time.

Figure 1 demonstrates how movement and rotation cues and precues update in a subtask and after the user finishes a subtask. Figure 1(a) shows the phase in which the user moves their hand to the object. A yellow line guides the user to the object. In Figure 1(b), the user is ready to pick up the object, causing their hand to become loaded. In Figure 1(c), the user transports the object while rotating it. In Figure 1(d), the user puts down the object, unloads their hand, and completes the subtask. This causes all line and wedge-set visualizations to brighten and thicken, and the first line and wedge set to disappear. Meanwhile, a new dimmer and thinner line and dimmer and smaller wedge set are added.

The number of line and wedge-set visualizations can be different. In the rest of this paper, we use uLvW to denote them, where u is the number of cyan lines (L) used, while v is the number of wedge sets (W) used, noting that each wedge set contains a green wedge and a gray wedge. We do not include the yellow line since it does not provide enough information to finish the movement part of a subtask.

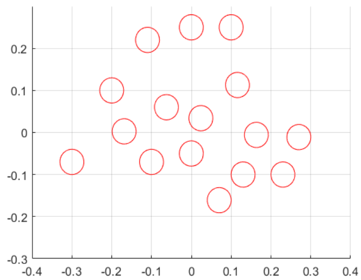

Fig. 3. Distribution of the 16 targets in our test environment. In each of the 16-subtask-long test sequences used in the studies, all targets are visited once. The horizontal and vertical axes represent x and z coordinates on the table used in the studies, measured in meters.

4 AR Testbed

We implemented an AR testbed to help investigate the potential benefits and limitations of movement and rotation precues [21]. The user interface and associated application interface were developed using Unity 2019.4.18f1 [40] and the Mercury Messaging framework [9].

4.1 Test Scene

The current test scene (Figure 3) consists of sixteen targets, each of which can serve as the source location for an object or the destination to which it is to be moved in a subtask. Each 3cm-radius target is represented by a red circle, which is close to the radius of the base of the Vive Tracker 2.0 units used as the objects. Targets are distributed on a 50cm×80cm area on a circular table parallel to the xz-plane. We made sure that seated users could reach all targets. In Figure 3, all targets are visualized. However, in our user studies, only the cued and precued targets are visualized to prevent participants from memorizing test sequences. We decided to test 1–4 (cyan) lines and 1–4 wedge sets. Five objects are used, so there will be an object that is not precued even in the condition with a cued subtask and three precued subtasks.

4.2 Sequences

There are 16 subtasks in each of the sequences we use. Each subtask includes the four phases described in Section 3.1. Each subtask was designed such that no object was precued twice at any time. In the 16 subtasks, each object was used in three to five subtasks. The sequences were designed to form a closed loop (in terms of the distribution of objects in the scene), so at the end of the task, the distribution of objects will be the same as at the beginning. This means we can pick a random start point in the sequence while making sure all tasks using that path traverse the same set of segments.

The distance between an object and its destination was either 10cm, 15cm, 20cm, or 25cm. Each appeared four times in a 16-subtask task.

In each subtask, the user was asked to rotate the object by one of the following angles around the y-axis: $-144°$, $-72°$, $+72°$, $+144°$. Each of these four angles appears four times in a sequence. Subtask difficulty was determined by movement distance, movement direction, and rotation angle, rather than only one of these.

Since, at the beginning of the trial, the five objects might not be at the correct positions in the sequence, we insert five setup subtasks to make the participant move them to the correct starting targets for each trial. The lines and wedge sets work the same way in the setup subtasks as in the 16 regular subtasks. The only difference is that the objects might not be on the targets. The setup subtasks were excluded from the analysis.

To avoid confounding effects related to a participant's anticipation of completing a trial, we present an entire set of additional subtasks, past the point of the last subtask to be completed, such that our precueing system will render additional visualizations past the end. Thus, unless a participant is counting the subtasks, which we found to not be the case in piloting, they will not anticipate the end of a trial.

5 User Study One

We divided User Study 1 into two parts. In Part 1, we investigated the impact of different wedge-set configurations. In Part 2, we investigated the best number of lines and wedge sets.

5.1 Hypotheses

Based on our initial design goals, and observations and results from our pilot studies, we formulated three hypotheses. For Part 1, we hypothesize:

H1. *Participants will spend less time finishing a subtask with the O/D wedge-set configuration than with the O/O or D/D configuration.* Based on our pilot studies, we expected that the O/D wedge-set configuration would be better than O/O and D/D, since it allows the user to work on rotation and movement at the same time.

For Part 2, we hypothesize:

H2. *Compared to the condition that only uses one line and one wedge set to cue movement and rotation for the current subtask, adding another line or another wedge set will improve subtask completion time.* We believe precueing the movement or rotation information of the next subtask will help the user prepare for it, improving performance. Moreover, in the study by Liu et al. [20], participants performed best with two to three places precued with line visualizations. As mentioned in Section 3.3, this corresponds to our condition in which two lines are used.

H3. *Performance will deteriorate when more than some optimal number of precues are used.* We believe that this will occur because the scene will become too cluttered for the task to be performed as effectively as when fewer precues are used.

5.2 Methods

5.2.1 Participants

Our study was approved by the review board of our institution. We recruited 29 participants from our institution (11 female), 18–29 years old (average 23.2), five left-handed, through convenience sampling accomplished using email sent through our department email lists and posted flyers. Among the participants, 17 had no AR/VR experience (six finished only Part 1), eight had used AR/VR several times (one finished only Part 1), three had used VR in class projects, and one owned a VR headset. Each participant was compensated with a \$15 gift card. Participants were recruited in two phases: 24 in the first phase, and five in the second phase. Of the participants recruited in the first phase, two did not complete Part 2. Participants in the second phase were recruited for the analysis in Section 6.2 and only did Part 1. Thus, 29 participants completed Part 1, and 22 completed Part 2.

5.2.2 Equipment

Study participants wore a Varjo XR-3 video–see-through AR headset [41] with a 115° horizontal field of view (134° diagonal at 12 mm eye relief) and a 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. During the study, they were allowed to use only their dominant hand. Five Vive Tracker 2.0 units were used as tracked objects. All trackers were tracked by the same base stations that tracked the XR-3. The XR-3 ran on a computer powered by an Intel® Core™ i9-11900K Processor and an Nvidia GeForce RTX 3090 graphics card. The trackers were put on a 46-cm radius circular. The headset, trackers, and table were sanitized using 70% isopropanol before each session.

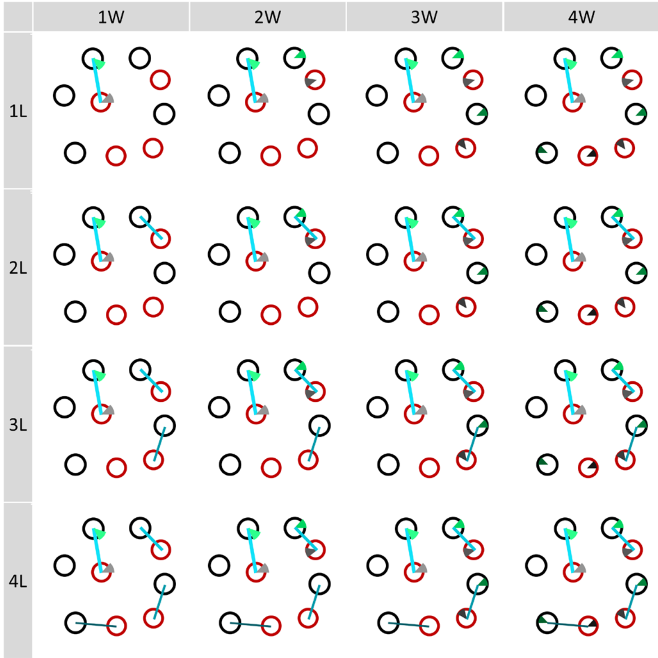

Fig. 4. The 16 conditions tested in Study 1 Part 2. All conditions use the O/D configuration. Black and red circles represent trackers and their destinations.

5.2.3 Study Design

Trials were blocked by condition. Each block consisted of one untimed practice trial and two timed trials. Each trial contains a 16-subtask sequence (task). The order in which the blocks appeared was shuffled in a counterbalanced manner based on the condition to which it belonged, ensuring each participant experienced a different order of conditions.

Part 1 uses 12 conditions: (O/O, O/D, D/D)×(1L1W, 2L2W, 3L3W, 4L4W). The aim of Study 1 Part 1 was to find the best wedge-set configuration, so we blocked the conditions by number of precues rather than wedge-set configuration, counterbalancing block order and wedge-set configuration order within blocks. We did this to avoid any participant encountering a specific wedge-set configuration only at the beginning, middle, or end of Part 1. Part 2 uses 16 conditions: (1L, 2L, 3L, 4L)×(1W, 2W, 3W, 4W). For Part 2, we use the O/D configuration because it outperformed the other two configurations in our pilot studies. Schematics of conditions tested in Part 2 are shown in Figure 4.

5.2.4 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) [34] to screen for stereo vision and the Ishihara Pseudo-Isochromatic Plate (PIP) test [14] to screen for color deficiencies. All participants passed the SFT and PIP tests. The study coordinator then put the headset on the participant and attached the Vive Tracker 3.0 to the participant's dominant hand using a Rebuff Reality TrackStrap. The study coordinator then calibrated the XR-3 eye tracker and 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 formal trial blocks.

Throughout the study, data from the headset and trackers, the trial and subtask completion times, and the participant's eye-gaze data 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, a modified unweighted NASA TLX [26], and a request to rank the techniques based on their effectiveness. The TLX survey was modified to use a 1–7 scale, with 1 as best rather than the original 0–20. Each participant rated each of the three wedge-set configurations for each TLX metric. Images of each wedge-set configuration were displayed during the rating process to remind participants of the visualizations used in the three configurations. As we mentioned in Section 5.2.3, we blocked the conditions by number of precues, so there was no way to interleave within Part 1 three separate TLXs, one for each wedge-set configuration, such that each TLX would occur only after experiencing all trials with its wedge-set configuration. Instead, we chose to give only a single questionnaire at the end of the study. This also 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 one and a half to two hours for a typical participant to complete.

5.3 Results

We sought to understand if there was a relationship between the rotation/movement precue placement and user performance, measured in completion time. Note that a task could not be completed without completing each constituent subtask. This was verified by detecting collision of a small sphere collider (radius $= 0.1$cm) located at the center of the tracker with capsule colliders bounding the 3cm-radius targets and by checking whether the tracker orientation is within 15° of the specified orientation. We picked the same positional tolerance as Liu et al. [20] so we could compare how the added complexity in our task would affect the users' performance. Subtask completion time is computed from the time the collider of the last object collided with the collider of the corresponding destination to the time the collider of the current object collided with the collider of the current destination. Therefore, the time a user deposits an object will be included in the next subtask rather than the current subtask. We did this because it is hard to measure whether a user had deposited the object based on the hand-position information, as depositing might be done solely using fingers rather than the full hand. Section 6.1 will discuss the types and numbers of errors made in each condition.

On study completion, we processed completion-time results generated automatically by our system before analyzing them. We used Tukey's outlier filter [39] to label outliers before analysis. The "outside fence" for each condition and user was computed separately. We expected the conditions would have a significant effect on completion time, and we noticed that some users performed substantially better than others. Subtasks that took more than the third quartile plus 1.5 times interquartile range (third quartile minus first quartile) or less than the first quartile minus 1.5 times interquartile range were labeled as outliers. The percentage of subtasks labeled as outliers in conditions in Part 1 and Part 2 ranged from 2.56% to 5.40%. The percentage of subtasks labeled as outliers in each condition can be found in the supplementary material.

Average subtask completion time after outlier removal under each condition is shown in Figure 5. Different participants and subtasks performed differently. For example, our subtask movement lengths range from 10cm to 25cm, and we would expect a participant to take significantly less time to perform a 10cm movement than a 25cm movement. In addition, each subtask differs in movement direction and rotation angle. Therefore, rather than directly calculating the standard error of all subtasks, we used linear mixed-effects models to calculate the contribution of each factor. (See the linear mixed-effects model results in the supplementary material for the standard error.)

We evaluated all hypotheses for significance with $\alpha = .05$. For each hypothesis, we fit a separate linear mixed-effects model to our data using the MATLAB Statistics and Machine Learning Toolbox [22].

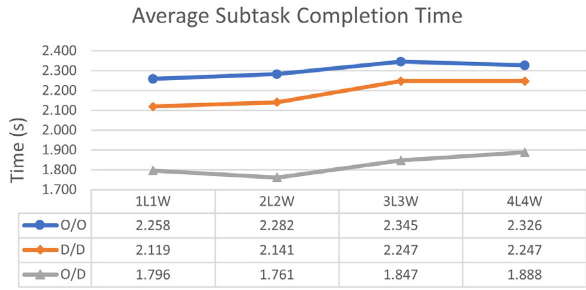

Fig. 5. Study 1 Part 1 subtask completion time. O/D performed best, followed by D/D. O/O performed worst.

Table 2 lists the parameters used in each model. **M1**–**M4** denote the models used to evaluate **H1**–**H4**, respectively. Each model used the subtask completion time as the measurement. We used subtask (the specific subtask in the 16-subtask sequence) and user as the random-effect variables. There is no interaction term between random-effect variables. By comparing different linear mixed-effects models, we found that handedness and AR/VR experience are not significant factors. Color-vision deficiency, in Study 2 is also not a significant factor. These may be because they were absorbed by the user random effect.

H1 was evaluated using **M1**. To make evaluation of **H1** easier, we used O/D-2L2W as the baseline condition. These terms are picked based on comparing the current model and other alternative models using a likelihood ratio test. The p-values for the wedge-set configuration fixed-effect terms are $< .0001$. The 1L1W, 3L3W, and 4L4W terms have a p-value of .33, .003 and $< .0001$, respectively. On the other hand, the p-values for the interaction terms are all $> .05$. This shows that the effect of wedge-set configuration is significant, and the effect of the number of precued subtasks on the increase of completion time after three subtasks cued and precued is also significant. The interaction between them is rather weak. For the full details of the linear mixed-effects model, please see the supplementary materials.

To evaluate **H1**, we examined the p-values of the fixed-effect terms for rotation visualization configuration. Since we used O/D as the baseline condition, we checked the p-values for the O/O and D/D fixed-effect variables. Both are $< .0001$, supporting that O/D performs better than O/O and D/D.

For Part 2, the average subtask completion time after outlier removal under each condition is shown in Figure 6. The standard error contributed by each factor, analyzed by the linear mixed-effects models, is shown in the supplementary material together with the linear mixed-effects model results. Here, the average subtask completion time does not have a clear trend, but we noticed that 1L2W and 2L1W performed better than the other conditions. This indicates that the performance might be related to the total number of visualizations (L+W) in the scene. Figure 7 plots the average segment completion time against the total number of visualizations in the scene. It can be seen that three visualizations yield the best completion time, followed by two and four visualizations. Participants performed worse in the other conditions.

We used **M2** to evaluate **H2**. We use one line and one wedge set as the baseline condition. To evaluate **H2**, we compared two pairs: 2L1W vs. 1L1W and 1L2W vs. 1L1W. For each pair, the p-values from the linear mixed-effects model are .0005 and .0076, respectively. This shows that participants did better with 2L1W and 1L2W than with 1L1W, supporting **H2**.

We used **M3** to evaluate **H3**. The standard error contributed by each factor, analyzed by **M3**, is shown in the supplementary material together with the linear mixed-effects model results. Three total visualizations are used as the default condition. In order to check if this model fits the data well, we compare **M3** with **M2**. A likelihood ratio test shows that **M3** contains slightly less information than **M2**, but the p-values for the fixed-effect terms are significant and smaller ($< .0001$). This shows that the total number of cues is a good descriptor for the performance. Since

Table 2. Parameters of the linear mixed-effects models.

Model	Data Source	Fixed-effect variable(s)	Interaction terms	Effect Sizes
M1	Study 1 Part 1	Wedge-set configuration, Number of precued subtasks	Yes	$\eta^2 = 0.5870$, Cohen's $d = 1.0167$ (Large)
M2	Study 1 Part 2	Number of lines, Number of wedge sets	Yes	$\eta^2 = 0.6039$, Cohen's $d = 1.0460$ (Large)
M3	Study 1 Part 2	Total number of visualizations	N/A (Single variable)	$\eta^2 = 0.7619$, Cohen's $d = 1.3196$ (Large)
M4	Study 2	Number of lines, R/NoR	No	$\eta^2 = 0.7441$, Cohen's $d = 1.2889$ (Large)

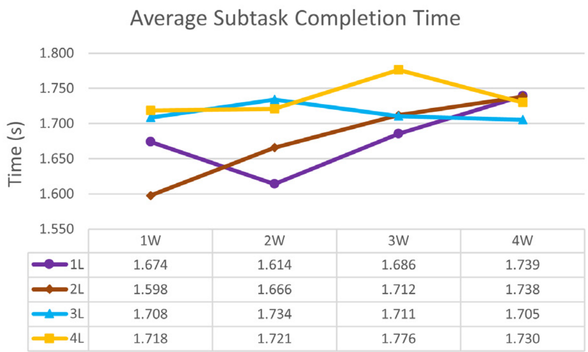

Fig. 6. Average participant performance (seconds per subtask) plotted by number of lines and number of wedge sets. Participants performed best with 1L2W and 2L1W.

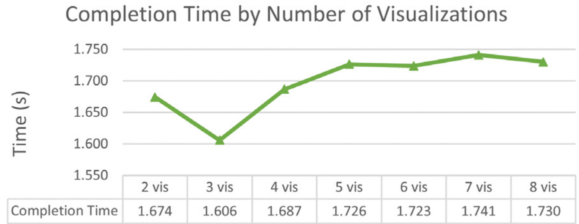

Fig. 7. Average participant performance (seconds per subtask) plotted by total number of visualizations. Participants performed best with three visualizations; performance degraded when more visualizations were used.

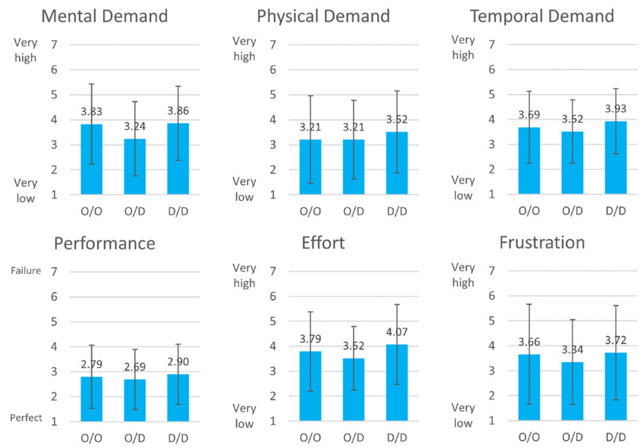

Fig. 8. NASA TLX results. Error bars indicate the standard deviations. Generally speaking, we found no significant difference between the three placements.

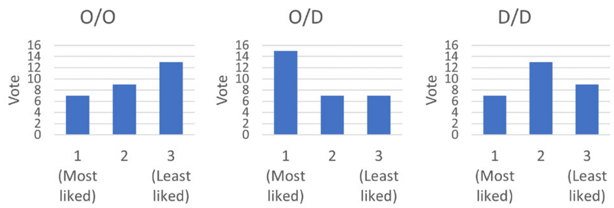

Fig. 9. Histograms of user preferences for wedge-set configuration. Generally speaking, O/D is most preferred, while O/O is least preferred.

three total visualizations is the baseline condition, the conditions with more than three visualizations all performed worse, and the p-values for the comparisons are all significant ($< .0001$), **H3** is supported.

To avoid type-I errors, we ran a correction using the Holm–Bonferroni method [13]. We checked a total of 10 p-values in the linear mixed-effects models (2 for **H1**, 2 for **H2**, and 6 for **H3**) to validate our hypotheses. Among the ten p-values, eight of them are $< .0001$, one is .0005, and one is .0076 (lowest-to-highest). With this order, the p-values are smaller than $.05/10$, $.05/9$, ..., $.05/1$, respectively, meaning all tests survive their corresponding Holm–Bonferroni-corrected α.

5.3.1 User Feedback

At the end of the session, each participant was given a questionnaire that included demographic information (including which equipment they used), a set of custom unweighted NASA TLX surveys, and questions on user preference.

We first let the participants answer a separate modified unweighted NASA TLX for each wedge-set configuration in the questionnaire. The results are shown in Figure 8. We conducted Friedman tests on the NASA TLX results, yielding $p_{MentalDemand} = .0263$, $p_{PhysicalDemand} = .0090$, $p_{TemporalDemand} = .0329$, $p_{Performance} = .3614$, $p_{Effort} = .1897$, and $p_{Frustration} = .5709$. The p-values for performance, effort, and frustration are $> .05$. The p-values for mental demand, temporal demand, and physical demand are also not significant with Holm–Bonferroni correction. Thus we did not find a significant difference between the techniques on these NASA-TLX metrics. However, as shown in the linear mixed-effects model used to evaluate **H1**, the users objectively performed the best in O/D, in terms of completion time.

Participants then ranked wedge-set configurations by their preferences. The results are shown in Figure 9. Similar to the NASA TLX results, none of the three placements significantly wins out. We ran a Friedman test for the preference result, showing the difference between the three is insignificant ($p_{Preference} = .1664$).

Participants were also asked how many lines and wedge sets they thought were best for their performance. For the number of lines, five participants preferred one, fourteen preferred two, three preferred three, and none preferred four. For the number of wedge sets, eleven participants preferred one, eight preferred two, three preferred three, and none preferred four. Most users preferred two lines, which is consistent with their having performed the best in 2L1W. On the other hand, half the users preferred one wedge set over two wedge sets, which is not consistent with their having performed the best in 1L2W.

6 Tracking Data Analysis

6.1 Error Analysis

To understand the kinds of errors participants made in Study 1, we looked into the subtasks labeled as outliers with Tukey's method [39]. Since errors could have multiple causes, we decided to label them manually. We found that participants made the following major types

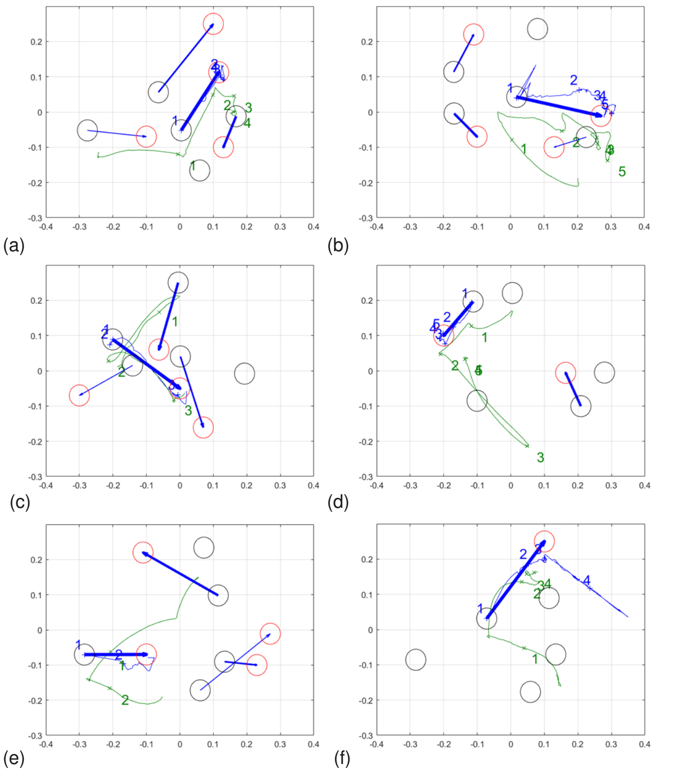

Fig. 10. Examples of common error types in outlier subtasks. (a) Rotation Adjustment. (b) Movement Adjustment. (c) Wrong Target. (d) Incomplete. (e) Fast. (f) Tracking. Blue and green paths represent the current object and hand traces, respectively. The horizontal and vertical axes represent *x* and *z* coordinates on the table used in the study, measured in meters. Numbers on the green and blue paths increment with elapsed time and can be used to correlate hand and object motions.

of errors in Study 1:

Rotation Adjustment (1.29%): In these cases, the participant picked up the tracker but spent a relatively long time adjusting the tracker's orientation. As shown in Figure 10(a), the tracker did collide with the destination, but the subtask was not finished since the participant needed to adjust the orientation. Sometimes the participant adjusted the orientation right after picking up the tracker rather than after arriving at the destination.

Movement Adjustment (0.62%): In contrast to the Rotation Adjustment error, in movement error, the participant spent time adjusting the position of the tracker. As shown in 10(b), in this kind of error, the participant typically moved the tracker to a place near the destination and pushed it back to the destination after they realized the movement part had not been finished.

Wrong Target (0.84%): As shown in Figure 10(c), the participant moved their hand toward an incorrect tracker. In some cases, they picked up the correct tracker but moved toward an incorrect destination. They moved to the cued destination in rare cases before picking up the tracker. Only in very rare cases did the participants go to the destination the cyan line connected rather than the tracker, which may be because the red circle at the destination and the Vive Tracker are sufficiently different so the user would not be confused by them.

Incomplete (0.10%): As shown in Figure 10(d), the participant moved to the next tracker without finishing the current one. After they realized that, they moved back and finished the current subtask.

Fast (0.07%): Few cases were labeled as outliers because the participant spent a very short time finishing them. These normally involve short paths, but there is nothing special in the tracker's path. Figure 10(e) shows one example.

Tracking (0.35%): In rare cases, tracking would fail and cause the users to spend additional time on some segments. This is typically indicated by unreasonably quick tracker movement or a discrepancy between the hand path and the tracker path, as shown in Figure 10(f).

Others (0.62%): All other errors that are not covered by the previous error types. Most of these involve a combination of the other six types. In some other cases, the participants spent a long time planning the movement, but no anomalies were found in the hand or tracker paths.

The percentages of each error type under each condition can be found in the supplementary material. We used chi-square tests to analyze if the participants made more specific types of errors under specific conditions. All comparisons either have an insignificant p-value ($< .05$) or have a small effect size ($\phi < 0.1$). This is because the percentage of each type of error is fairly low. The error analysis shows that participants' performance in each condition was reflected in their completion time rather than their error rate.

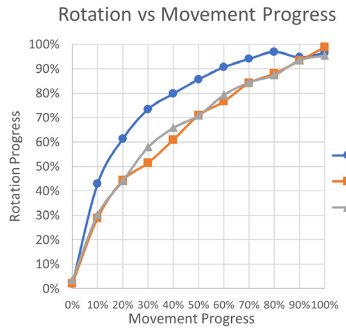

Fig. 11. Participants' rotation progress plotted against movement. Note that the lines do not need to be monotonic, since we recorded the rotation progress when the participant reached a specific movement-progress threshold.

6.2 Movement and Rotation Progress Analysis

In Figure 11, we plotted rotation progress against movement progress for the five participants recruited in the second phase, who only did Part 1 of the study. Generally speaking, the participants made progress on rotation earlier than on movement. This is consistent with the work by Novick and Tversky [27], which showed that participants preferred to work on rotation earlier than movement. For the trend in each of the three conditions, it can be seen that, for O/O, participants' rotation progressed faster than in the other two conditions. For D/D, the participant focuses more on the movement task. However, notice that the two axes in (a) are in percent, as the rotation and movement required for each subtask were different.

6.2.1 Discussion

Figure 5 shows that participants did best with O/D, followed by D/D, and then O/O. This means that for O/O, while the participants started tackling the rotation subtask earlier, they were slow in the movement, and therefore performed worse than O/D and D/D. Similarly, although both in D/D and O/D, the participants performed the movement task fast, they spent more time in D/D rotating the tracker than in O/D when the tracker is about to reach or is already on the destination. Therefore participants did worse in D/D than in O/D. To understand these trends better, we look into the participants' gaze traces.

Note that the data in Figure 11 is an average and individual strategies may differ, though each of the 29 participants in Study 1 Part 1 performed the best with O/D.

6.3 Gaze Analysis

We checked eye-gaze traces from the same five participants whose data were analyzed in Section 6.2 to get a sense of how participants are affected by the visualization. Figure 12 (a–c) shows several examples in the 2L2W condition. In these examples, when the participant's hand was unloaded, their eyes moved first while the hand followed. After

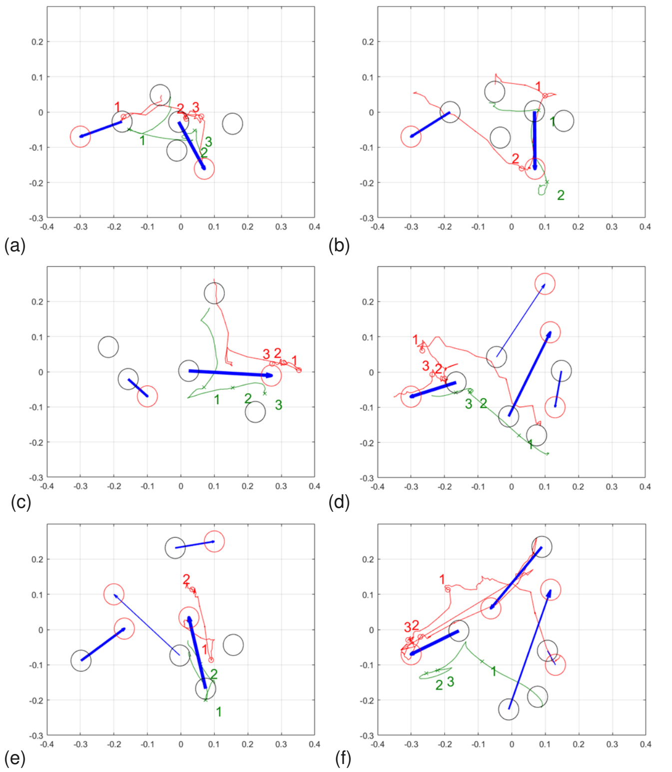

Fig. 12. Sample eye-gaze traces for different wedge-set configurations and different numbers of lines and wedge sets. The horizontal and vertical axes represent x and z coordinates on the table used, measured in meters. (a) O/O-2L2W. (b) O/D-2L2W. (c) D/D-2L2W. (d) O/O-4L4W. (e) O/D-4L4W. (f) D/D-4L4W. Black circles represent the objects (Vive trackers). Red circles represent the targets to which the objects should be moved. Cyan arrows are the line visualizations showing the movement information. Red and green paths represent the participant's eye gaze and hand traces, respectively. Numbers on the red and green paths increment with elapsed time and can be used to correlate eye and hand motions.

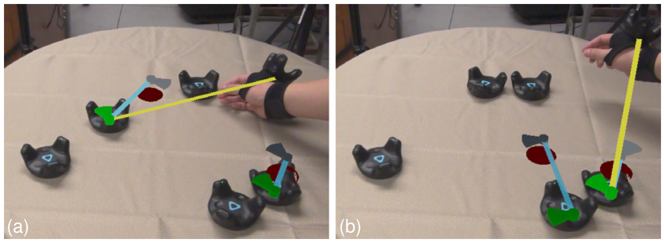

Fig. 13. Subsequent subtasks in the 2L2W-NoR condition. (a) The first wedge set on the upper left shows correct information for the current subtask, while the second wedge set on the lower right does not show rotation for the next subtask. Its gray wedge shows the current tracker orientation rather than the goal orientation. (b) After the participant finished the first subtask in (a), the lower right subtask becomes the current subtask, and its gray wedge switches to show the goal orientation. Hence, the wedge set gave the participant rotation information. A second wedge set that did not show the rotation information for the next subtask appeared.

loading their hand, the eye-gaze trace differs based on the conditions (the rotation visualizations). Figure 12 (d–f) shows several examples of eye gaze in the 4L4W condition.

6.3.1 Impact of Wedge-Set Configurations

In O/O (Figure 12a), the participant's eye gaze stayed on the tracker and did not move to the destination before the loaded hand and the tracker were transported. This is consistent with the result shown in Figure 11. Participants were able to perform the rotation earlier, perhaps because a larger portion of the visualization (cues and precues) fell on the trackers rather the destinations. However, not focusing on the destination caused participants to spend more time on the movement. In O/D and D/D (Figures 12b and c), a larger portion of the visualization fell on the destinations rather than the trackers. This prompted participants to look more at the destinations, and since the participants could rely on proprioception to know where the current tracker was, they did not need to switch their gaze between the tracker and the destination very often. The participant's eye gaze moved faster to the destination than the hand and the object after the object was picked up. This is also consistent with the result in Figure 11. The participant focused first on the movement and then the rotation. Note that the participants' gaze stayed on the destination longer in D/D (Figure 12c) than in O/D (Figure 12b). This can be explained by the fact that they spend more time on the destination to perform the rotation in D/D, and is also consistent with the result in Figure 11. Comparing O/D and D/D, the participant spent more time in D/D rotating the tracker than in O/D when the tracker was about to reach or was already on the destination. The difference can be explained by the different natures in O/D and D/D.

6.3.2 Impact of Number of Visualizations

Study 1 Parts 1 and 2 both show that two lines and wedge sets led to worse performance. We examined the participant's gaze trace in order to understand the cause of the trend. In Figure 12 (e), there was no other precue on the way from the last destination to the current object, so the participant's eye gaze went directly to the object. In Figures 12 (d) and (f), however, there were other visualizations near the cue or on the participant's way from the last destination to the current object. The participant's eye gaze was attracted by them and did not directly go to the current object. This suggests that too many lines and wedge sets cause worse performance because the participants would see them and need to decide whether they were the current cue. Even in Figure 12(e), the participant still performed more slowly than in the condition with fewer lines and wedge sets. Further, human ability to visually track objects is limited [1, 6, 24, 30]. When too many lines and wedge sets are used, participants might lose track of the order of the lines and the wedge sets, and therefore perform worse.

6.3.3 How Much Rotation Information Can Participants Use?

In Part 2 of Study 1, participants performed best in 1L2W and in 2L1W. One question here is whether the movement precue or the rotation precue is more important. As discussed in Section 5.3.1, a majority of participants reported they felt two lines were the most useful. When asked about wedge sets, half thought one wedge set was best and half thought two or three was best. This raises the issue of whether participants could use the precued rotation information. Moreover, in the examples in Figure 12, we notice participants need to look at the tracker to work on the rotation. However, even in the cases of 2L2W in Figure 12(a–c), the gaze did not stay on the second line and wedge set. Looking at Part 2, we used the O/D wedge-set configuration. In this configuration, the two wedges show the positions of the object and the destination, which means they also contain the movement information of the subtask. The participant could use them to finish the movement part of a subtask by moving the green wedge to where the gray wedge is located. To check whether, in the 1L2W condition, the participant used the movement information rather than the rotation information of the second wedge, we decided to conduct a follow-up study—User Study 2.

7 User Study Two

The recruiting process, equipment, design, and study procedure were the same as described in Section 5, with the exception of participants and conditions tested. We recruited eight participants from our institution: three female, 21–25 years old ($\bar{X} = 23.4$), one left-handed.

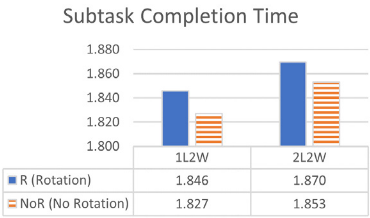

Fig. 14. Study 2 subtask completion time. Eliminating rotation information did not cause a significant change in performance.

Among the participants, five had no AR/VR experience, two had used AR/VR several times, one had used VR in class projects, and none owned a VR headset. One has a color-vision deficiency and could not answer any question in the PIP test correctly. We formulated one hypothesis:

H4. *The participant's performance will not degrade in 1L2W and in 2L2W if we eliminate the rotation information in the second wedge set.* We expect participants to use the movement information rather than the rotation information of the second wedge set, so eliminating its rotation information will not negatively affect performance.

Study 2 used four conditions, all with the O/D wedge-set configuration: (1L2W, 2L2W)×(R, NoR), where R and NoR denote the second wedge set has and does not have rotation information, respectively. In NoR, the second gray wedge did not show rotation information for the subtask. We accomplished this by making the second gray wedge show the tracker orientation rather than the target orientation. When a participant finished a task and the second (precueing) gray wedge became the first (cueing) gray wedge, its orientation would switch to the target orientation of the current task. By doing this, we eliminated the rotation information of the second wedge set, but the participants were still able to use it as a movement precue, as it indicated both the tracker and the destination positions of the first upcoming task. An example of the NoR condition is shown in Figure 13.

We wanted to know whether the relative performance of 1L2W and 2L2W (with R) is consistent with Study 1 Part 2, which had different participants.

7.1 Results

The percentage of subtasks labeled as outliers using Tukey's outlier filter ranged from 2.34 to 6.25%. The detailed percentages for each condition can be found in the supplementary material. Figure 14 shows the subtask completion time. We ran a linear mixed-effects model, **M4**, to evaluate **H4**. The parameters of **M4** can be found in Table 2. The standard error contributed by each factor, analyzed by **M4**, is shown in the supplementary material together with the linear mixed-effects model results. We compared **M4** with an alternative linear mixed-effects model with interaction terms between the number of lines and R/NoR fixed-effect variables. The interaction was neither significant nor strong, so we decided to use the model without the interaction term. We found that adding R/NoR as a fixed-effect variable will not make the model contain more information and it will not have a significant p-value (.4094). To evaluate **H4**, we ran a non-inferiority test on the R/NoR term and picked a non-inferiority margin of 0.05s (completion time increased by 0.05s). The p-value of the test is .007, showing eliminating the rotation information in the second wedge set will not degrade the performance under the non-inferiority margin. Therefore, **H4** is supported.

In addition, seven of the eight participants indicated in the questionnaire and exit interview that they did not notice when the second wedge set did not show correct rotation information. (This may be explained by change blindness [31].) Therefore, we conclude that participants could not use the rotation information in the second wedge set. They could use the movement information of two subtasks (four places cued or precued) and the rotation information of one subtask. We analyzed the percentages of the error types (defined in Section 6.1) the participants made in Study 2 and found no significant difference between each condition. The percentages of each type can be found in the supplementary material.

Looking back to H2 in Section 5, although it is still true that participants did better in 1L2W than in 1L1W, the results of Study 2 show that it was because participants used the movement information in the second wedge set rather than its rotation information. This means that participants in our studies could not use the precued rotation information in the second wedge set.

8 Discussion

8.1 Comparison with Work on Path-Following Task

In our studies, participants performed best when provided movement information for two subtasks cued and precued. As mentioned in Section 3.3, two places were visited in each subtask. One place was cued, and depending on whether the participant's hand was loaded, there were either two or three places precued. Thus, our results are consistent with Liu et al. [20], which showed that participants could make use of two to three precues in their VR path-following task. However, the best number of places precued might be related to how smooth the path is. In a smooth path such as the one used by Ellis et al. [8], the way that a place precue is defined might be very different, causing the result to change.

8.2 Can Participants Use Rotation Precues?

Our studies show that participants were able to use rotation information only for the current subtask. This may be because they needed to also work on the movement part of the subtasks and keep moving their head and eyes. In Section 6.3, we saw that participants needed to look at the wedge-set visualization that showed the rotation information in order to work on it. If the visualizations that cue and precue rotation information stayed on fixed positions (no movement is needed in such a case), and potentially in the participant's foveal vision, it may be possible to utilize one or more rotation precues. This may also be related to whether the participant is trained. For example, DJs are able to manipulate multiple knobs on their mixers simultaneously. Since they are trained and familiar with the layout of the controls, it might be possible for them to parse one or more rotation precues.

8.3 Field of View and Rotation Visualization

In our testbed, the movement distance for each subtask is at most 25cm. When a participant looks at a object or a destination, the other (destination or object) will also fall in their field of view, so that they can use the line and the wedge set simultaneously. However, if the object and the destination are far enough apart that the participant will not be able to see one of them when they look at the other, the best visualization might change. In that case, the movement and rotation visualization can be merged as a path that indicates how the participant should pre-rotate their hand to pick up and drop off the item. For example, one can use visualizations like the Attention Funnel [4], but connecting the user's hand to the destination with the correct orientation, or a ParaFrustum [36].

8.4 Design Guidelines for Precueing Visualizations

Through our user studies, we learned that among the wedge-set configurations, it was more useful to split the rotation information between an object and its destination. Besides, in addition to the movement and rotation information of the current subtask, the participant was able to parse the movement information of the next subtask, which means they can parse the information for four places at the same time. Combining these, we infer that when designing precueing visualizations for manual tasks, one should not overemphasize either the object or the destination in the rotation information, as it would potentially decrease performance. Besides, participants' ability to parse rotation information is rather limited compared to movement information. Therefore, for precueing, one might want to focus on movement information more. For visualizations that cue rotation information, one can even try to embed the movement information, just like the O/D configuration. A participant's performance tends to degrade when the scene contains too much clutter, so one might want to eliminate unnecessary visualizations. However, we expect that the number of visualizations participants can utilize is task-dependent and additional research will be necessary to find the cut-off point.

9 Limitations and Future Work

9.1 Participant Population and Sample Size

As mentioned in Section 5.2.1, we recruited the participants from our institution and all were relatively young. Further study will be needed to determine if our results generalize to older people.

Another limitation of our work is the relatively small number of participants in the studies. Although we were able to get significant p-values to evaluate the hypotheses in Study 1 with 29 participants and in Study 2 with eight participants, we note these groups do not necessarily reflect the performance of the general population.

9.2 Tasks Requiring Specific Rotational Directions

In our task, the users only needed to rotate items to the specified orientation and the direction to rotate was not specified. This is true of many real-life tasks, such as orienting objects in CAD systems and during hardware assembly. However, there are other tasks in which an object must be rotated in a specific direction. For example, when fastening/loosening a bolt, the driver must be rotated clockwise/counterclockwise. These cases could be addressed by using 3D arrows.

9.3 Precueing Using Eye-Gaze Information

Section 6.3 implies that performance may be related to how cluttered the scene is, or, more specifically, how many visualizations are shown in the participant's field of view and the region between the cueing visualization and the first precueing visualization. It is worth investigating whether dynamically adjusting the number of precues based on how cluttered the user's field of view is could help improve performance. Another possibility is, like Wolf et al. [44], to show a warning if the user is moving in the wrong direction.

9.4 Precueing Parallel and Collaborative Subtasks

The task discussed in this paper is a single-user, serial task. However, some tasks require the user to perform parallel subtasks. For example, in *osu!* [7], the user may need to accomplish subtasks for the left hand and right hand at the same time, as in playing most musical instruments. In some cases (e.g., *Overcooked 2* [37]), multiple users can address the same set of subtasks. In these situations, precueing subtasks might improve performance. However, showing parallel precues might also increase a player's cognitive load and negatively affect their performance. It will be important to investigate how to precue parallel and collaborative subtasks.

9.5 Precueing for Body Navigation

Our study used cues and precues to guide arm movements. It should also be possible to precue movements of the entire body, for example, in navigation [25,38]. In some cases, accomplishing a task like equipment maintenance and repair requires a user to first move their entire body to the right location and then use head, eyes, arms, and hands [36]. Precueing should help performance in such tasks, although the mode and timing of the precues are likely to be different from those used here.

9.6 Precueing for VR Mid-Air Manipulation

We address manipulation on the $xz-$plane. However, other work, such as that by Vuibert et al. [43] and by Mendes et al. [23], focuses on mid-air object manipulation in VR. In such cases, the manipulation becomes 6DoF and the best cueing/precueing approach might be different.

10 Conclusions

We explored how graphical cues and precues could be used to provide information about future subtasks in AR for a heterogeneous manual task that requires the user to move their empty hand to one of a set of objects, pick it up, move it to its destination while rotating it, and then put it down. Pilot studies favored a line visualization for precueing movement information and a wedge-set visualization for precueing rotation information. Our first user study tested different configurations for the wedge-set visualizations and found that participants performed best when rotation information was split between the manipulated object and its destination. In addition, participants performed best when given two lines and one wedge set or one line and two wedge sets, that is, three pieces of information including at least one line and one wedge set. Our second study showed that eliminating the rotation information of the second wedge set did not reduce performance significantly.

Therefore, we conclude that the participants performed best when given the movement information of the current subtask and one upcoming subtask (four places total) and the rotation information for only the current subtask. This stands to reason, as translation seems to be a fast movement of the arm and shoulder to a destination whereas rotation requires fine adjustment of the wrist and fingers. Holding the position of the wrist and fingers while making a large fast movement of the entire arm might be difficult, especially for novices. We believe our results could be applied to a wide range of real-world tasks, including maintenance and assembly tasks, which require moving and rotating a series of target objects.

Acknowledgments

This research was funded in part by National Science Foundation Grant CMMI-2037101. We thank Portia Wang for her assistance in making the video.

References

[1] G. A. Alvarez and S. L. Franconeri. How many objects can you track?: Evidence for a resource-limited attentive tracking mechanism. *Journal of Vision*, 7(13):14–14, 10 2007. doi: 10.1167/7.13.14

[2] S. Andrist, M. Gleicher, and B. Mutlu. Looking coordinated: Bidirectional gaze mechanisms for collaborative interaction with virtual characters. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, p. 2571–2582. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3025453.3026033

[3] 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

[4] F. Biocca, A. Tang, C. Owen, and F. Xiao. Attention funnel: Omnidirectional 3D cursor for mobile augmented reality platforms. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '06, pp. 1115–1122. ACM, New York, NY, USA, 2006. doi: 10.1145/1124772.1124939

[5] L. Chen, Y. Liu, Y. Li, L. Yu, B. Gao, M. Caon, Y. Yue, and H.-N. Liang. Effect of visual cues on pointing tasks in co-located augmented reality collaboration. In *Symposium on Spatial User Interaction*, SUI '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3485279.3485297

[6] N. Cowan, C. Morey, and Z. Chen. The legend of the magical number seven. In S. D. Sala, ed., *Tall tales about the brain: Separating fact from fiction*, pp. 45–59. Oxford University Press, 2007.

[7] Dean Herbert. *osu!* `https://osu.ppy.sh/home`, last accessed on 03/10/2022.

[8] S. R. Ellis, F. Breant, B. Manges, R. Jacoby, and B. D. Adelstein. Factors influencing operator interaction with virtual objects viewed via head-mounted see-through displays: Viewing conditions and rendering latency. In *Proceedings of IEEE 1997 Annual International Symposium on Virtual Reality*, pp. 138–145. IEEE, 1997.

[9] 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

[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] M. Hertzum and K. Hornbæk. The effect of target precuing on pointing with mouse and touchpad. *International Journal of Human–Computer Interaction*, 29(5):338–350, 2013. doi: 10.1080/10447318.2012.711704

[12] K. Higuch, R. Yonetani, and Y. Sato. Can eye help you? effects of visualizing eye fixations on remote collaboration scenarios for physical tasks. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems*, CHI '16, p. 5180–5190. Association for Computing Machinery, New York, NY, USA, 2016. doi: 10.1145/2858036.2858438

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

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

[15] F. Jeanne, I. Thouvenin, and A. Lenglet. 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*, VRST '17. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3139131.3139144

[16] A. Jing, K. May, G. Lee, and M. Billinghurst. Eye see what you see: Exploring how bi-directional augmented reality gaze visualisation influences co-located symmetric collaboration. *Frontiers in Virtual Reality*, 2:79, 2021. doi: 10.3389/frvir.2021.697367

[17] G. Klinker, O. Creighton, A. Dutoit, R. Kobylinski, C. Vilsmeier, and B. Brugge. Augmented maintenance of powerplants: a prototyping case study of a mobile ar system. In *Proceedings IEEE and ACM International Symposium on Augmented Reality*, pp. 124–133, 2001. doi: 10.1109/ISAR.2001.970522

[18] F. Kosmalla, F. Daiber, F. Wiehr, and A. Krüger. Climbvis: Investigating in-situ visualizations for understanding climbing movements by demonstration. In *Proceedings of the 2017 ACM International Conference on Interactive Surfaces and Spaces*, ISS '17, p. 270–279. Association for Computing Machinery, New York, NY, USA, 2017. doi: 10.1145/3132272.3134119

[19] K. S. Lashley. The problem of serial order in behavior. In L. A. Jeffress, ed., *Cerebral mechanisms in behavior*, vol. 21, pp. 112–136. Bobbs-Merrill Oxford, United Kingdom, 1951.

[20] 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

[21] J.-S. Liu, B. Tversky, and S. Feiner. A testbed for exploring multi-level precueing in augmented reality. In *2022 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW)*, pp. 540–541, 2022. doi: 10.1109/VRW55335.2022.00121

[22] The MathWorks, Inc. *Matlab Statistics and Machine Learning Toolbox*. `https://www.mathworks.com/help/stats/index.html`, last accessed on 03/10/2022.

[23] D. Mendes, F. Fonseca, B. Araùjo, A. Ferreira, and J. Jorge. Mid-air interactions above stereoscopic interactive tables. In *2014 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 3–10, 2014. doi: 10.1109/3DUI.2014.6798833

[24] G. A. Miller. The magical number seven, plus or minus two: Some limits on our capacity for processing information. *Psychological review*, 63(2):81, 1956.

[25] W. Narzt, G. Pomberger, A. Ferscha, D. Kolb, R. Müller, J. Wieghardt, H. Hörtner, and C. Lindinger. Augmented reality navigation systems. *Universal Access in the Information Society*, 4(3):177–187, 2006.

[26] NASA. Task load index (tlx) v. 1.0 manual. *Human Performance Research Group, NASA-Ames Research Center Moffett Field*, 1986.

[27] L. R. Novick and B. Tversky. Cognitive constraints on ordering operations: The case of geometric analogies. *Journal of Experimental Psychology: General*, 116(1):50–67, 1987.

[28] 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

[29] A. C. Pierno, C. Becchio, M. B. Wall, A. T. Smith, L. Turella, and U. Castiello. When Gaze Turns into Grasp. *Journal of Cognitive Neuroscience*, 18(12):2130–2137, 11 2006. doi: 10.1162/jocn.2006.18.12.2130

[30] Z. W. Pylyshyn and R. W. Storm. Tracking multiple independent targets: Evidence for a parallel tracking mechanism. *Spatial vision*, 3(3):179–197, 1988.

[31] R. A. Rensink, J. K. O'Regan, and J. J. Clark. To see or not to see: The need for attention to perceive changes in scenes. *Psychological Science*, 8(5):368–373, 1997. doi: 10.1111/j.1467-9280.1997.tb00427.x

[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] 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

[34] Stereo Optical Co. Inc. *Original Stereo Fly Stereotest*. `https://www.stereooptical.com/products/stereotests-color-tests/original-stereo-fly/`, last accessed on 03/10/2022.

[35] 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

[36] M. Sukan, C. Elvezio, O. Oda, S. Feiner, and B. Tversky. Parafrustum: Visualization techniques for guiding a user to a constrained set of viewing positions and orientations. In *Proceedings of the 27th Annual ACM Symposium on User Interface Software and Technology*, UIST '14, p. 331–340. Association for Computing Machinery, New York, NY, USA, 2014. doi: 10.1145/2642918.2647417

[37] Team17 and Ghost Town Games. *Overcooked 2*. `http://www.ghosttowngames.com/overcooked-2/`, last accessed on 03/10/2022.

[38] B. Thomas, V. Demczuk, W. Piekarski, D. Hepworth, and B. Gunther. A wearable computer system with augmented reality to support terrestrial navigation. In *Digest of Papers. Second International Symposium on Wearable Computers (Cat. No.98EX215)*, pp. 168–171, 1998. doi: 10.1109/ISWC.1998.729549

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

[40] Unity. *Unity*. `https://unity.com/`, last accessed on 03/10/2022.

[41] Varjo. *Varjo XR-3*. `https://varjo.com/products/xr-3/`, last accessed on 03/10/2022.

[42] 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

[43] V. Vuibert, W. Stuerzlinger, and J. R. Cooperstock. Evaluation of docking task performance using mid-air interaction techniques. In *Proceedings of the 3rd ACM Symposium on Spatial User Interaction*, SUI '15, p. 44–52. Association for Computing Machinery, New York, NY, USA, 2015. doi: 10.1145/2788940.2788950

[44] J. Wolf, Q. Lohmeyer, C. Holz, and M. Meboldt. Gaze comes in handy: Predicting and preventing erroneous hand actions in ar-supported manual tasks. In *2021 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, 2021.

[45] 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