Multi-Level Precues for Guiding Tasks Within and Between Workspaces in Spatial Augmented Reality

Benjamin Volmer, Jen-Shuo Liu, Brandon Matthews, Ina Bornkessel-Schlesewsky, Steven Feiner and Bruce H. Thomas

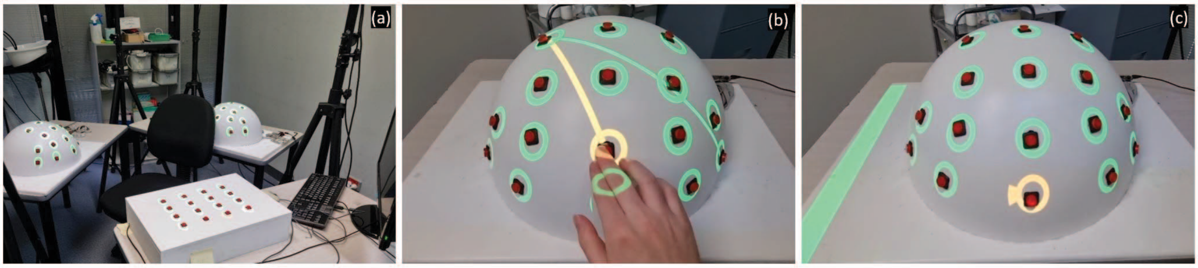


Fig. 1: Three workspaces located around the user's chair with projected annotations to guide the user. (a) Study environment. (b) Workspace annotated with two LINE precues to guide the user within the workspace. (c) Workspace annotated with an ARC precue, including a workspace-shift precue (green bar) that indicates the upcoming task will be in the workspace to the left.

Abstract— We explore Spatial Augmented Reality (SAR) precues (predictive cues) for procedural tasks within and between workspaces and for visualizing multiple upcoming steps in advance. We designed precues based on several factors: cue type, color transparency, and multi-level (number of precues). Precues were evaluated in a procedural task requiring the user to press buttons in three surrounding workspaces. Participants performed fastest in conditions where tasks were linked with line cues with different levels of color transparency. Precue performance was also affected by whether the next task was in the same workspace or a different one.

Index Terms—Spatial augmented reality, predictive cues, precueing, cognitive load, multiple tasks, procedural task

1 Introduction

Procedural tasks are intrinsic to numerous fields of work [13]. Research has shown that different forms of Augmented Reality (AR) [9, 15], such as Spatial Augmented Reality (SAR) [8, 37], which provides hands-free viewing, can improve performance during procedural tasks. AR can overlay virtual instructions on or near tasks, allowing for visually seamless guidance through a sequence of operations in domains as diverse as medicine [1, 16, 29, 32], manufacturing [10, 12], assembly [40, 46], maintenance [14, 21, 49], and control [5, 24, 27].

Recent research has shown that visually indicating the next location using a precue can further improve performance beyond cueing the current location, reducing the mental effort for completing AR-enhanced procedural tasks [43, 44]. This aligns with prior findings that a precue can improve reaction times when using a mouse or touchpad and a 2D display [19]. Research has also demonstrated that precueing additional tasks beyond the upcoming one can improve performance even further in headset-based Virtual Reality (VR) [25] and AR [26]. This leads us to pose the research question: *Does providing multiple precues improve performance for SAR-enhanced procedural tasks?*

Furthermore, studies investigating VR and AR precues have focused on performing tasks in a single workspace [25, 26, 44]. In contrast, some real-world scenarios require users to shift between multiple workspaces, such as a pilot interacting with multiple control panels [48] and a nuclear power plant operator managing systems in a control room [23]. These observations lead to our second research question: *Can SAR precues improve performance when shifting between workspaces is required?*

To answer these questions, we designed an experiment that required users to press buttons in a specified order in three different workspaces located around them, as shown in Fig. 1(a). We employed heterogeneous configurations of button devices (two curved domes and one flat panel). Our work expands upon existing research on precueing to make the following three contributions:

- We investigated multiple levels of precues in SAR. Previous work in SAR has investigated only a single level of precueing.
- We conducted the first investigation of precues in multiple workspaces, including shifting between them. Multiple workspaces correspond to important domains such as control rooms and aircraft cockpits.
- We created and evaluated new precue visualizations that combine cue type and color to support switching between workspaces.

2 Previous Work

2.1 Augmented Reality and Procedural Tasks

As previously noted, AR allows the virtual and real worlds to be co-located [2, 7, 35]. For example, in procedural tasks, virtual instructions can be integrated with a workspace, rather than requiring separate paper or electronic documentation [39]. Events in the task domain can also be tracked and interactively reflected in the virtual content [50].

Assembly tasks require users to follow instructions. Wu et al. [47] compared instructions for building a toy presented by paper and AR. The paper condition required users to read instructions to determine the next part's position and orientation. The AR instructions highlighted the part and provided the correct placement details in situ. Participants strongly preferred the AR instructions, as they simplified the task. Tang et al. [40] evaluated three presentation conditions (monitor, paper-

- *Benjamin Volmer, Brandon Matthews, Ina Bornkessel-Schlesewsky and Bruce H. Thomas are with the University of South Australia. E-mail: { benjamin.volmer | brandon.matthews}@mymail.unisa.edu.au and { bruce.thomas | ina.bornkessel-schlesewsky}@unisa.edu.au .*
- *Jen-Shuo Liu and Steven Feiner are with Columbia University. E-mail: jl5004@columbia.edu and feiner@cs.columbia.edu.*

Manuscript received 25 March 2023; revised 17 June 2023; accepted 7 July 2023. Date of publication 24 October 2023; date of current version 31 October 2023. This article has supplementary downloadable material available at https://doi.org/10.1109/TVCG.2023.3320246, provided by the authors.
Digital Object Identifier no. 10.1109/TVCG.2023.3320246

based, and AR guidance), determining that AR guidance improved task completion time and accuracy compared to the other conditions.

In medicine, AR can be used to co-locate information with a patient's body during a procedure [29]. For example, AR has been employed to display the distance between tumors and blood vessels [16] and indicate locations at which to insert an implant [22] or give an injection [17].

AR instructions have improved performance and user experience for interacting with procedural tasks. Marner et al. [27] compared traditional monitor-based instructions with SAR-projected instructions for pressing buttons on two different devices in separate conditions: a dome and a car dashboard. Results revealed SAR improves response time and causes fewer errors. Baumeister et al. [5] continued this work by observing the monitor and dome throughout a sleep deprivation study in which participants stayed awake for one night. Results indicated participants in the monitor condition suffered from a speed–accuracy tradeoff, while those in the projected AR condition maintained a similar level of performance. Baumeister et al. [6] then examined SAR instructions compared to a monitor and two AR head-worn displays (optical-see-through [11] and video-see-through [34]). They used a dual-task methodology [33] that required users to press a button to their side when prompted by auditory stimuli. For the primary button-pressing task on the dome and attending the secondary task, SAR annotations resulted in the fastest response times.

2.2 Studies of AR and VR Precues

Recent research in AR and VR has shown that presenting predictive information for the upcoming task location can further improve performance. Volmer et al. [44] conducted a series of experiments utilizing a dome control panel for which they first designed five SAR precues to guide users to a single upcoming task. From these cues, linking a line between tasks was shown to provide superior response times, accuracy, and lower self-assessed mental effort. Volmer et al. then investigated the same set of precues in a sleep deprivation study [43] and in a study using electroencephalography (EEG) [36] to measure electrical brain activity [45]. Observations from these two studies indicated that using a precue can offset the negative effects of sleep deprivation and that some types of precues can even reduce the cognitive resources required for a task, evaluated by brain-response data.

Liu et al. [25] investigated multiple levels of precues in VR. Similar to Volmer et al. [44], they discovered that visualizations with lines were the best. In addition, they found that participants performed the best with two or three levels of precues when the visualizations included lines and with one level of precue when the visualizations did not include lines. Subsequently, Liu et al. [26] explored multiple levels of precueing in AR, addressing tasks in which a user must move and rotate each of a series of objects, one at a time, placing it at a designated position and orientation. They found that in these more complicated tasks, participants could use one movement precue but not a rotation precue.

Building on this work, we apply multi-level precueing techniques to a SAR procedural task. Furthermore, unlike prior research, rather than addressing just a single workspace, we explore precues within and between multiple workspaces surrounding the user to better approximate complex real-world environments.

3 Study Conditions

The main task investigated in this study involved pressing buttons in three workspaces located around the user, shown in Fig. 1(a). This configuration was chosen to provide an extreme case of workspace placement, where some tasks are out of the participants' direct line of sight. Each button was surrounded by a circle SAR annotation. SAR precues guided the participant to the next button or group of buttons. These precues moved to the next button of a sequence as the task progressed. This continued until all buttons in each workspace were pressed correctly once.

3.1 Precues

We designed precues based on three factors: Cue (NONE, ARC, and LINE), Color (SOLID and TRANSPARENT), and Multi-level (1, 2, and 3). These cues were inspired by earlier research on attention directional cues when working with large 2D monitors [20], projected links showing information movement between desktop computers [38], and visualizations of the locations of offscreen data [4].

The Cue factor is the visual annotation displayed in addition to the circle annotations. Three styles were measured, each of which provides a different type of information about the step. The NONE style (Fig. 2a) provides only *place* information of the destination in the step, without an ARC or LINE annotation, and this leaves the users to rely on the colored circle annotations. ARC provides *direction* information by attaching an annotation to the active or upcoming task (button), pointing in one of eight compass directions towards the next button in the sequence (Fig. 2b). Finally, LINE provides *path* information via linking a line between the current and upcoming buttons (Fig. 2c).

The Color factor comprises two different color formats. SOLID (SOL) only modifies the active button's annotation color to orange. This includes the ARC or LINE cue attached to it. Any upcoming precues remain green to maintain focus on the primary task, and to prevent confusion caused by multiple precues being the same solid shade of orange. TRANSPARENT (TRN) is the other condition, which modifies all precue annotations up to the Multi-level (described below) to become different shades of orange. The further into the sequence, the more transparent the annotations become. Comparing Fig. 2c and Fig. 3 shows the differences between SOL and TRN for the LINE condition.

The final factor is Multi-level (Multi), which has three levels: 1, 2 and 3. This factor determines how many annotations will appear to guide users to the active and upcoming tasks. Level 1 modifies the active task based on the prior Cue and Color factors. Level 2 provides an additional modified precue at the upcoming button location. The ARC and LINE cues linking from the second to the third task are thinner, and the annotations' color is more transparent for the TRN conditions. Multi-level 3 continues this pattern, further reducing the ARC and LINE annotation sizes and making the TRN conditions more transparent. Fig. 2 shows how the ARC and LINE cues become smaller for the higher level of Multi and how TRN annotations become darker in the NONE condition.

Table 1: Precues tested in our study. Conditions crossed out are duplicate visualizations that were removed.

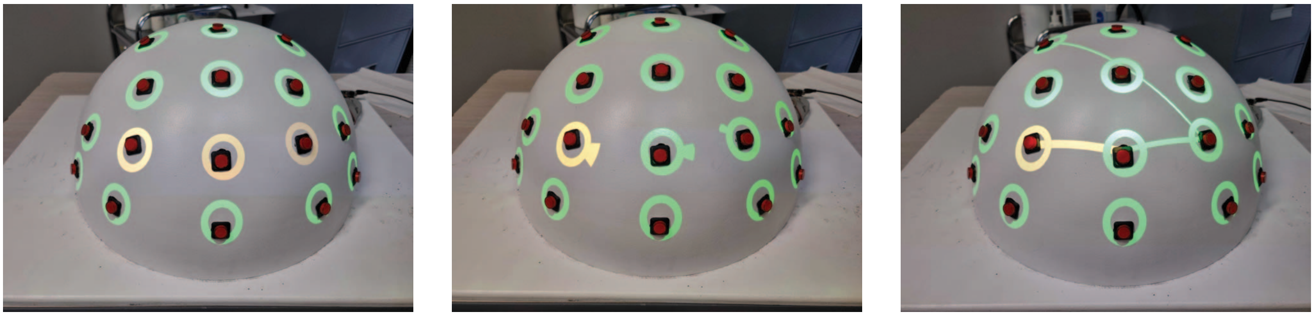
	1		2		3	
NONE	TRN	~~SOL~~	TRN	~~SOL~~	TRN	~~SOL~~
ARC	TRN	~~SOL~~	TRN	SOL	TRN	SOL
LINE	TRN	~~SOL~~	TRN	SOL	TRN	SOL

Each precue we designed is a combination of all three factors, leading to a 3×2×3 experimental design. However, some factor combinations visually appear the same. NONE TRN 1 and all Multi-levels of NONE SOL would change only the color of the active task. Thus, the three Multi-level factors of NONE SOL were not evaluated. Furthermore, Multi-level 1 for the LINE and ARC conditions were also not differentiated by the Color factor. As Multi-level 1 modifies only the active task location, SOL and TRN were visually the same. Therefore, only TRN was evaluated for Multi-level 1 conditions. Thus, five conditions were dropped to avoid repeating the same visualization, leaving the study with 13 unique conditions, shown in Tab. 1.

3.2 Workspace-Shift Cues

In addition to pressing buttons in the same workspace, users sometimes had to rotate to another workspace roughly 120° to their left or right, as shown in Fig. 1(a). To help determine which way the user should rotate, all cue conditions provide additional guidance. We designed three workspace-shift cues, each of which provides *place*, *direction*, and *path* about the upcoming workspace, respectively. These match the information provided by the NONE, ARC, and LINE styles.

NONE (Fig. 4a) highlights the upcoming workspace orange, allowing users to notice a visual change in their peripheral vision. The



(a) NONE: No ARC or LINE annotation is provided. The active button is highlighted in orange. Depending on the Multi-level (in this case 3), the upcoming one or two buttons to press are also highlighted successively more transparent shades of orange.

(b) ARC: An arc is attached to the active task annotation, pointing in the compass direction of the upcoming task. Depending on the Multi-level (in this case 3), additional arcs are attached to the upcoming conditions, becoming thinner the later they are in the sequence.

(c) LINE: A line annotation is linked between the active and upcoming task. Depending on the Multi-level (in this case 3), additional successively thinner lines appear later in the sequence.

Fig. 2: Same-workspace precues. The conditions are shown with a Multi-level factor of 3. This means that all conditions will have three cues, each successively thinner or more transparent the later it is in the sequence.

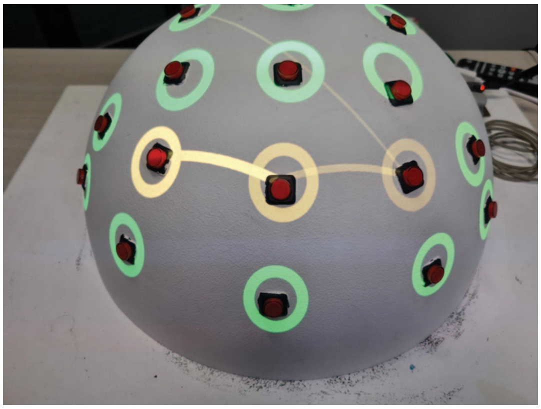


Fig. 3: LINE TRN Multi 3 condition. Differs from LINE SOL Multi 3 shown in Fig. 2c as the additional line annotations are different colors. Further in the sequence, the lines become thinner and more transparent.

upcoming device was modified as there is no directional cue factor to help guide the user from the active device to the upcoming one.

ARC (Fig. 4b) has the annotation on the final task of the active device point left or right towards the direction of the upcoming workspace. An additional green bar annotation is also projected on the side of the active workspace. The bar is colored green to avoid having an additional orange augmentation in the current workspace that is not attached to the button annotations. The bar annotation further emphasizes that the user would have to rotate to another device for the upcoming button. Without the bar annotation, the user might interpret the ARC pointing left or right as indicating the button next to the active task is the next button to press, not a button on another device.

LINE (Fig. 4c) links the line annotation in the direction of the upcoming workspace. These additional cues appear based on the number of remaining button presses in the active workspace and the Multi-level value. For example, at level 1, the visualization appears when the last button to be pressed on the active workspace is highlighted. Once the button is pressed, the additional annotation vanishes, and the user rotates based on the predictive information provided by the precue. For Multi 2 and 3, the annotation appears when two or three presses remain in the active workspace, informing users they have to rotate after two or three correct presses.

3.3 Pilot Studies

We conducted several pilot studies to determine the precue conditions described in the prior section. An initial pilot study with four participants tested directional annotations without changing the color of the annotation alongside versions in which the active task was a different color or blinking. It evaluated various combinations of existing precues from the work conducted by Volmer et al. [44] and investigated performance in a more complex environment of multiple workspaces. Results from this pilot demonstrated that a blinking active task was too distracting. Additionally, highlighting the active task in a different color improved switching between workspaces.

We then tested changing the active annotation's color in combination with a direction annotation and two and three levels of precues. This pilot study also examined a condition that used different shades of the same color and one that used entirely different colors. Performance was better using the same color with varying transparency. Participants noted a lower level of mental effort for different shades of one color, as they did not have to learn color schemes. For these reasons, we used one color with varying levels of transparency for this research.

We performed an additional pilot study to examine Multi-levels of precues between one and four. Aligning with work by Liu et al. [25], four levels of precues caused performance to decay heavily. Thus, we decided to use three levels of precueing in our main study to reduce the total number of comparisons.

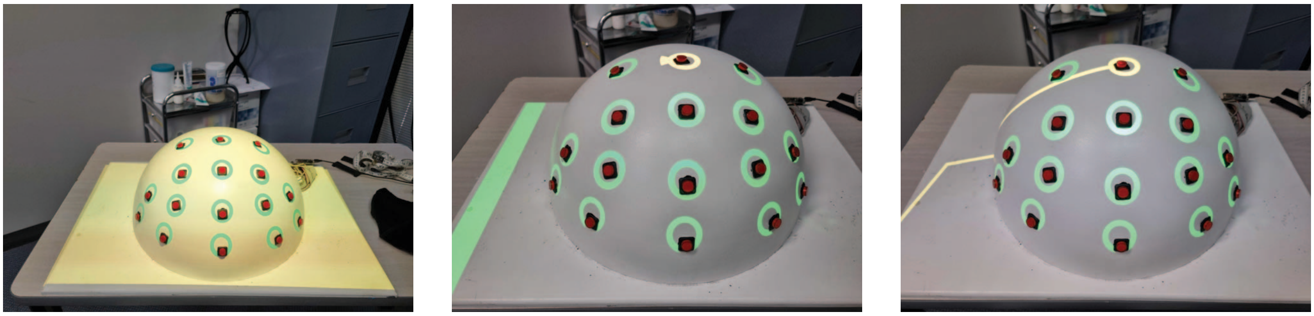
4 Experimental Design

This research was conducted to investigate if SAR precues can be beneficial in a complex work environment with multiple workspaces. We first asked whether multiple precues can be applied to within-workspace tasks (H1) and workspace-shift tasks (H2). Then, we examined which type of visualization works best for single precue conditions (H3) and multiple precue conditions (H4). Moreover, our pilot study showed that modifying the directional cues to have different color transparencies could further improve performance (H5). Finally, we expected the type of visualization to impact user behavior in workspace-shift tasks (H6). We formulated the following hypotheses:

H1. *More precues will enhance performance when working in the same workspace.* This is inspired by Liu et al. [25], which showed that using multiple precues can improve user performance. We would like to determine if this can carry over to a SAR environment.

H2. *More precues will decrease performance when shifting to another workspace.* When switching to a different workspace, the user will encounter all precues in that workspace for the first time, since it was previously in the user's far peripheral vision. Thus, using more precues will increase the effort and time needed after the switch to distinguish the current task from those that follow.

H3. *Combining an arc or line precue with color will further enhance performance compared to just using color with no directional annotation.* Previous work investigated precues that either highlight the next task in a different color or show the direction to the next task. Com-



(a) NONE: The upcoming workspace becomes highlighted by an orange color, seen in peripheral vision.

(b) ARC: A green annotation appears on the left or right side of the active workspace.

(c) LINE: The line extends off the current workspace in the direction of the upcoming workspace.

Fig. 4: Workspace-shift precues. The cues appear X (the Multi-level of precue) levels in advance when the final tasks on the active workspace are being performed. These annotations are designed to guide to the upcoming workspace.

bining color and direction provides the user with two complementary ways to distinguish between the current task and next task.

H4. *Conditions with line annotations will perform better than other cues when working in the same workspace or shifting to a different workspace.* Previous work [25, 44] showed that visualizations with lines can improve user performance more than ones without lines.

H5. *Multiple precues with different shades of color will be superior to multiple precues with just the active task being a different color.* Our initial pilot studies showed that modifying the directional cues to have different shades of the same color improved performance.

H6. *Highlighting the position of the upcoming workspace will cause the user to move their head and hand toward that location more efficiently.* Since the user can see the other workspaces in their peripheral vision, the noticeable change of color towards their left or right will assist them in rotating in the correct direction. This is inspired by the work of Bailey et al. [3].

4.1 Participants

Adults 18–40 years old that did not have a color vision deficiency were eligible for this study. Due to a low room ceiling, participants were also required to be under six-feet tall. Twenty-five participants (19 male, 6 female) were recruited from our institution. Ages ranged from 19 to 39 ($M = 28.6$, $SD = 5.54$. Twenty-three participants were right-handed, while two were left-handed. Most participants stated they spend 40 or more hours a week using a computer ($M = 51$, $SD = 21.62$). Furthermore, 18 participants noted having prior experience with AR systems. Three participants were excluded because software or hardware errors occurred during their sessions, and one participant was excluded as an outlier for making too many errors, leaving twenty-one valid participants. This experiment was approved by the Human Research Ethics Committee of University of South Australia as a variation of approval number 200200.

4.2 Apparatus

Three workspaces were positioned around the user, as shown in Fig. 1(a). Two domes and one box, each with 16 buttons, were placed roughly 120° from each other on separate tables. If a participant looked directly at one workspace, they would see sections of both other workspaces in their far peripheral vision. A 1920×1080 monitor, connected to a nearby PC, was stationed behind the box device, and was used for completing questionnaires and instructing users which condition was next. Buttons were annotated by three Optoma EH500 DLP projectors, running at 1920×1080 (60Hz), stationed above or behind the workspaces. Each button device was connected to the PC through an Arduino HID device that treated each button press as a keystroke. The PC used an Nvidia GeForce GTX 1070 graphics card, an Intel i7 processor, and 16GB RAM.

Six OptiTrack Flex 13 cameras [30] were positioned around the room and used with Motive version 2.2 to track the user's head and hand via markers attached to a wristband and hat worn by the user. An application developed in Unity 2020.3.27f [42] connected to Motive to obtain head- and hand-tracking data for each frame. This application managed the button devices and played sounds for button presses. It also recorded response times (in milliseconds) between button presses and stored them in csv files. After each condition, the participant rated their mental effort between 1 (very, very low mental effort) and 9 (very, very high mental effort) using the Paas scale [31] on the PC.

4.3 Procedure

The participant was seated on a swivel chair in the middle of the three workspaces. They were first asked if they could rotate between each device comfortably. Participants were informed to use the index finger of their dominant hand and face the device on which they were working throughout the study. Once the participant was comfortable, they completed a short training session. This training session required them to press four buttons on two devices (8 total presses) for each of the 13 conditions. This made the participant aware of the annotation combinations and become used to shifting between workspaces. They were then shown an example scenario with all 48 (16×3) buttons annotated, to show the full set of buttons that would be used in the formal experiment. Participants did not interact with this demonstration.

The experiment began once the participant understood all 13 conditions. Each participant performed one block per condition, in fully randomized order. Each block contained three trials. One trial consisted of a 48-button press sequence. Each condition had three trials of different sequences to obtain a larger randomized dataset. To proceed through a trial, the correct button had to be pressed before the next precue would appear. Different sounds were played for correct and incorrect button presses. Between trials, all annotations would disappear, and the monitor would inform the user to press any button to start the next trial.

When the third trial of a condition was completed, the monitor would instead prompt the user to rate their mental effort. Each trial started in a random workspace.

The order of the buttons in a trial was randomized with a few exceptions. First, there would be eight workspace shifts per trial. Second, four to seven button presses happened in each workspace before a shift occurred. This allowed for the maximum Multi-level factor of three cues to appear on one device and prevented learning schemes by randomizing the number of same workspace button presses. The upcoming workspace for each shift was also randomized to avoid any learning patterns, such as constantly rotating to the left.

The total button-pressing time was approximately 35 minutes. After the study, a questionnaire was given to the participant to record their thoughts and preferences regarding the precue conditions. We followed a COVID-19 protocol that required the researcher to wear a mask and sit on the other side of the room from the participant. After each session, the three workspaces, wristband, hat, and PC input devices were sanitized.

5 Statistical Analysis

This section describes how the statistical analysis was conducted in our study. When analyzing performance-related data, we ignored the first button press of each 48-button press trial because the starting workspace was random and not precued. Participants had to search for the workspace in which it occurred, causing a longer response time. We also ignored the last press because it was not followed by a precue, changing its visualization.

Further cleaning of the performance-related data included removing three button locations. These included one button on the box that was intermittently unresponsive and one button on the corner of each dome because some participants remarked that they had difficulty differentiating colors projected on that button, apparently because of the curved surface and the projector's angle. The buttons whose presses were removed from the analysis for both domes were also discovered to have generated an outlier number of errors, supporting their removal from the analysis.

5.1 Response Time

Statistical analysis for response times was separated into two separate linear mixed-effects models (LMMs): one for pressing a button on the same workspace as the prior task and another for when the button was on a different workspace. Each block consisted of 144 correct button presses (3 trials $\times$ 3 workspaces $\times$ 16 buttons). Since each sequence required the user to shift workspaces 8 times, 24 presses were separated into the shift model per condition. Furthermore, 6 buttons were excluded for being the start or end of a trial leaving 114 for working on the same workspace. The number of buttons used for analysis fluctuated for each participant due to the removal of the three buttons highlighted in the prior section. Furthermore, correct presses over 5 seconds long on the same workspace were also removed. For shifting workspace, button presses over 6 seconds were removed. The extra second for changing workspace was to account for rotating to the new workspace. These long responses were removed to account for issues unrelated to the visual cues, for example, a phone call or when an Arduino or projector had connection issues. Fourteen presses were removed from working on the same workspace, and seven were removed for shifting workspaces. This left 28,665 correct presses on the same workspace and 6,173 for switching workspace.

Once data cleaning was complete, the two LMMs were created. Both models specified fixed effects of Cue, Color, and Multi-level alongside a random effect of the participant on the intercept. The significance of the main effects and interactions was assessed using Type-II Wald Chi-squared tests. Pairwise comparisons were then conducted using Tukey's honest significance test [41].

5.2 Errors

Akin to response times, errors were also separated into "same workspace" and "shift workspace" data sets. Due to the small number of errors across conditions, model observations could not be made effectively due to scaled gradient convergence issues. Visual observations from the figures are made instead. Additional comparisons are made between workspace error percentages. Chi-squared tests are used for comparing error percentages.

5.3 Head and Hand Tracking

We plotted the users' hand and head movements tracked with OptiTrack over time as moving traces using the data with Matlab R2020a [28]. By visually observing these traces, we categorized types of moving traces. In addition, to compare how the workspace-shift cues affect user behaviors, we compared users' hand and head movements over time under different conditions.

5.4 Self-Rated Mental Effort

The self-rated mental effort was evaluated using a model with fixed effects of Cue, Color, and Multi-level. A random effect of participants was also used. The main effect and interactions were analyzed using Type-II Wald Chi-squared tests followed by a post-hoc comparison utilizing Tukey HSD for corrections.

5.5 Preference

Separate figures and models were created for the participants' preference of each cue condition, multi-level and favored cue for working on the same or shifting workspace. A Kruskal–Wallis test was conducted on each preference-related response. This was followed by pairwise comparisons using Dunn–Bonferroni for corrections on comparisons.

6 Results

The results from our experiment are presented as follows: response time, errors, head and hand tracking, self-assessed mental effort and preferences.

6.1 Response Time

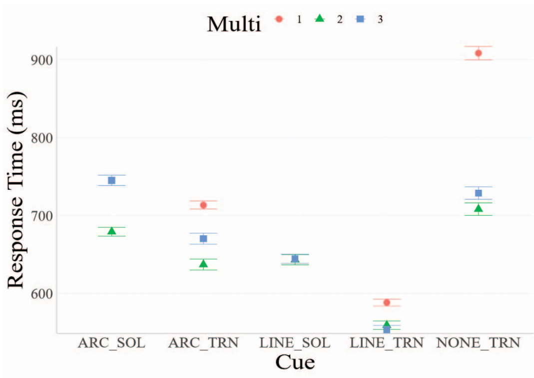


Fig. 5: Mean response times results for correct presses on the same workspace. Bars represent standard error.

For response times when working on the same workspace (Fig. 5), multiple significant main effects were discovered, showcased in Tab. 2. All main effects of Cue, Col, and Multi were significant alongside all two-way interactions. Pairwise comparisons also revealed many significant comparisons between the Cue, Col, and Multi factors (see supplementary material Table A). Some notable results included all conditions being significantly better than NONE level 1 and all LINE conditions being faster than all NONE conditions. Additionally, the level 2 and 3 conditions for ARC and LINE were significantly faster with multiple levels of color transparencies (TRN) compared to their single-color counterparts (SOL). Another interesting group of comparisons is between all LINE conditions. Comparing Multi-level 1 to the other levels (2 and 3), having multiple lines with varying levels of color transparency improves performance. However, Multi-level 2 and 3 for LINE annotations only modifying the color of the active task decreases response time.

Table 2: Type-II Wald Test results for response time comparisons of all conditions working on the same workspace. Significant values are shaded.

Condition	χ^2	df	p.value
Cue	1948.7586	2	$< .001$
Col	283.7958	1	$< .001$
Multi	509.4861	2	$< .001$
Cue:Col	10.8244	1	$< .002$
Cue:Multi	288.5692	4	$< .001$
Col:Multi	5.0702	1	$< .025$
Cue:Col:Multi	2.4150	1	$< .121$

Response times for shifting to another workspace are shown in Fig. 6. Similar significant main effects were found for response times when shifting to another workspace, see Tab. 3.

Post-hoc results (see supplementary material Table A) revealed many significant comparisons. As with the same workspace, there are too

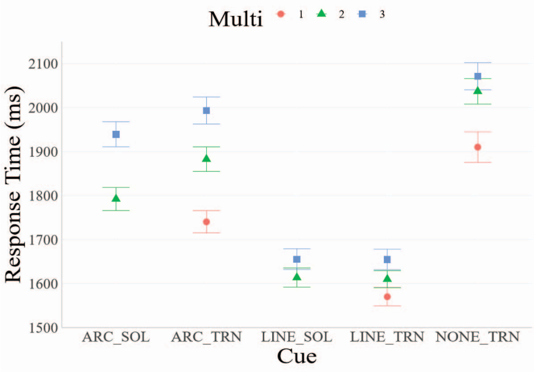


Fig. 6: Mean response times results for correct presses when switching to a different workspace. Bars represent standard error.

Table 3: Type-II Wald Test results for response time comparisons of all conditions when shifting to a different workspace. Significant values are shaded.

Condition	χ^2	df	p.value
Cue	548.2300	2	$< .001$
Col	4.9605	1	$< .026$
Multi	97.6800	2	$< .001$
Cue:Col	5.1942	1	$< .023$
Cue:Multi	20.5395	4	$< .001$
Col:Multi	.3418	1	$< .559$
Cue:Col:Multi	.3221	1	$< .571$

many comparisons to list so we focus on some of the particularly interesting pairs. All LINE precues were significantly faster than all other precues with the exception of the LINE 3 conditions in comparison to ARC 1.

When comparing the same precues with different levels, the level 1 conditions for ARC and NONE were significantly faster than their Multi 2 or 3 counterparts, with the exception of ARC 1 to ARC TRN 2. However, this was not the case for LINE. All LINE conditions were discovered to not have any significant differences when compared to each other. NONE also saw no difference between Multi-levels 2 and 3; however, both ARC conditions did.

Finally, neither LINE nor ARC saw a significant difference from the Color factors when comparing the Multi-level 2 and 3 conditions against each other. Response times for each workspace are shown in supplementary material Figs. A and B.

6.2 Errors

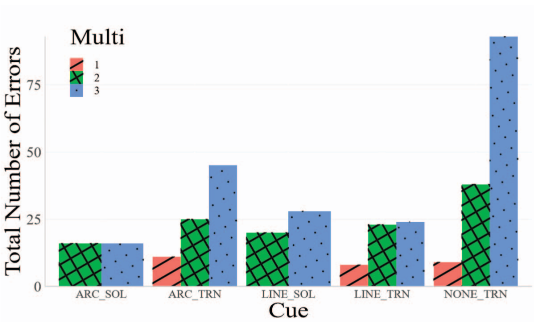


Fig. 7: Total number of incorrect button presses when working on the same workspace.

Due to the small number of errors, an LMM was unable to be successfully fit without convergence issues. Thus, we report error patterns based on visual inspection. Please note, however, that these should be interpreted with caution due to the lack of inferential statistics and the ceiling effects in regard to performance. When working on the same workspace (Fig. 7) it is apparent that presenting a higher number of precues causes a slight or large increase in the number of errors. ARC SOL appears to be least impacted by the increase in cues, followed by both LINE SOL and LINE TRN. NONE has the sharpest increase in errors when adding more levels of annotations. When observing the Color factor for ARC and LINE, SOL appears to be more accurate for ARC, while LINE shows little difference between SOL and TRN.

Table 4: Incorrect button presses made in each workspace.

	Box	Dome1	Dome2
Same Workspace	58 (0.616%)	127 (1.267%)	174 (1.734%)
Shift Workspace	13 (0.628%)	46 (2.197%)	44 (2.097%)

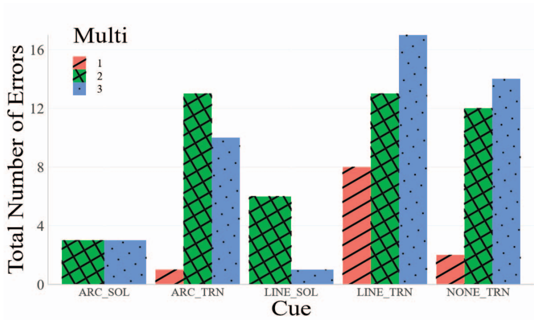


Fig. 8: Total number of incorrect button presses when switching workspace.

Even fewer errors occurred when shifting workspace (Fig. 8). With the current data set, it is hard to determine if any condition heavily impacts accuracy when shifting workspace; however, having a higher number of Multi-level precue with varying levels of color transparency may negatively impact performance.

As few errors occurred in individual conditions, we decided to investigate if particular devices may have impacted the number of errors. Observations revealed that the dome devices produced a higher percentage of errors, whether working in the same workspace or switching workspaces. Tab. 4 shows the percentage of incorrect vs. total presses per workspace. Chi-squared comparisons revealed that the percentages of errors on the box were lower than on the domes for same-workspace tasks ($p < .001$) and shift-workspace tasks ($p < .001$). However, the effect sizes are small ($\phi < 0.1$) because the error percentages are fairly low.

6.3 Head and Hand Tracking

By inspecting the plots of participants' hand-moving traces in within-workspace tasks, we found the traces could be categorized into three types: *Normal*, *Missed Button*, and *Wrong Button*. In normal, the participants moved their hands toward the correct button in the task. In *Missed Button*, the participants moved toward the correct button but failed to press it and moved their hands away from the button. Later on, they traveled back to finish the button pressing. We detected this in the tracking data if the participant moved their hand to within 5cm of the correct button and moved away to at least 10cm from that button without pressing the button. In *Wrong Button*, the participants moved toward the incorrect button and then, without hitting it, realized their mistake and turned to the correct one. We detected this if the participant moved their hand to within 5cm of a wrong button which is also not the goal of the last step or the current step for more than 15 frames in that step, but did not press the button. Tab. 5 lists the percentages of each trace type detected. We used chi-square tests to analyze if the percentages of the missed-button and wrong-button traces are higher

Table 5: Percentages of each trace type detected with tracking data in within-workspace tasks in each condition.

	ARC SOL	ARC TRN	LINE SOL	LINE TRN	NONE TRN
Missed Button	1.05%	0.88%	1.28%	1.54%	1.48%
Wrong Button	4.28%	4.11%	3.36%	5.62%	4.95%
Normal	94.67%	95.01%	95.36%	92.84%	93.57%

in some conditions. All comparisons either have an insignificant p-value ($> .05$) or have a small effect size ($\phi < 0.1$). This is because the difference between error rates in different conditions is fairly low.

We also observed how participants moved and rotated their hand and head in shift-workspace tasks to compare how workspace shift cues affected their performance. We compared the moving distance in different conditions and saw how participants performed over time. The result is shown in Fig. 9. Because different tasks involve different distances, instead of using absolute distance for the x-axes of Fig. 9, we normalized the tasks by discretizing the hand/head moving distance for each task in increments of 10% between starting hand/head position and ending hand/head position. We plotted the mean elapsed time across all tasks and participants at each of these 10% increments. The graph indicates that the participants' heads and hands moved the fastest in LINE, followed by ARC, and the slowest in NONE.

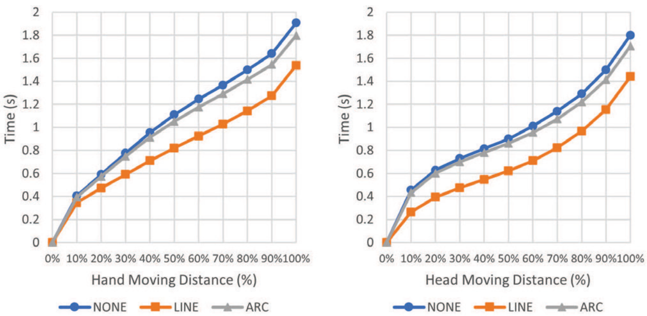


Fig. 9: Hand and head moving progress in workspace-shift tasks in mean elapsed time across all tasks versus normalized distance in percentage.

6.4 Self-Assessed Mental Effort

Two factors were found to be significant from the Chi-squared test on self-assessed mental effort. The main effects of Cue $\chi^2(2) = 75.9551, p < .001$ and Multi $\chi^2(2) = 7.3283, p < .026$. Mean results are visualized in Fig. 10. Pairwise comparisons also revealed a few significant differences listed in Tab. 6. The most notable being almost all LINE TRN conditions were significantly easier to process than all levels of the NONE condition. See supplementary material Table B for the full list of comparisons.

6.5 Preference

When asked about their preference for the five main Cue visualization styles (Fig. 11a), a significant main effect of Cue ($\chi^2(4) = 64.475, p < .001$) was observed. Dunn–Bonferroni post-hoc analysis revealed significant differences ($p < .001$) between LINE and NONE conditions. LINE SOL was also found to be superior ($p < .001$) to both ARC conditions. To a lesser extent, LINE TRN was also preferred over both ARC conditions ($p < .006$).

Three separate questions were also analyzed regarding how many Multi-levels of precue participants preferred for each Cue format (LINE, ARC, NONE), see Fig. 11b. From the Kruskal–Wallis tests, only NONE found a significant main effect of Multi ($\chi^2(2) = 10.685, p < .005$). Post-hoc tests only revealed observations for the NONE model. The

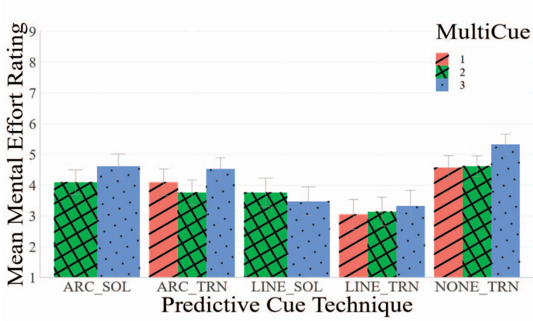


Fig. 10: Self-perceived mental effort results for each condition. Bars represent standard error. Ratings were between 1 (very, very low mental effort) and 9 (very, very high mental effort) akin to the Paas scale [31].

Table 6: PAAS mental effort scale ratings post-hoc for each condition. Only significant values are shown. Significant values are shaded.

Condition Comparison	estimate	SE	df	t.ratio	*p*.value
LINE TRN 1—NONE TRN 1	−1.5238	.361	265	−4.222	< .005
LINE TRN 1—NONE TRN 2	−1.5714	.361	265	−4.354	< .003
LINE TRN 1—ARC SOL 3	−1.5714	.361	265	−4.354	< .003
LINE TRN 1—ARC TRN 3	−1.4762	.361	265	−4.090	< .008
LINE TRN 1—NONE TRN 3	−2.2857	.361	265	−6.333	< .001
NONE TRN 1—LINE TRN 2	1.4286	.361	265	3.958	< .012
LINE SOL 2—NONE TRN 3	−1.5714	.361	265	−4.354	< .003
ARC TRN 2—NONE TRN 3	−1.5714	.361	265	−4.354	< .003
LINE TRN 2—NONE TRN 2	−1.4762	.361	265	−4.090	< .008
LINE TRN 2—ARC SOL 3	−1.4762	.361	265	−4.090	< .008
LINE TRN 2—ARC TRN 3	−1.3810	.361	265	−3.826	< .019
LINE TRN 2—NONE TRN 3	−2.1905	.361	265	−6.069	< .001
NONE TRN 2—LINE TRN 3	1.2857	.361	265	3.562	< .045
ARC SOL 3—LINE TRN 3	1.2857	.361	265	3.562	< .045
LINE SOL 3—NONE TRN 3	−1.8571	.361	265	−5.145	< .001
LINE TRN 3—NONE TRN 3	−2.0000	.361	265	−5.541	< .001

1 ($p < .009$) and 2 ($p < .027$) Multi-levels of Cue were preferred over having 3 levels of non-directional transparent annotations.

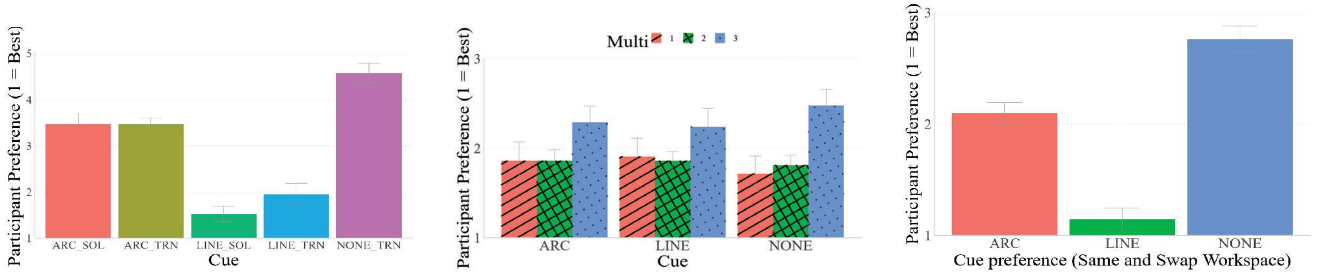
Finally, two separate questions were asked regarding the participants' favorite Cue style for working on the same workspace or shifting to a new one. Both data sets resulted in the same result, despite some participants listing different values for the separate questions. The mean results for both data sets can be observed in Fig. 11c. The Kruskal–Wallis test revealed Cue to be a significant effect ($\chi^2(2) = 41.052, p < .001$). Post-hoc comparisons further demonstrated that LINE was highly preferred ($p < .001$) over both ARC and NONE. ARC was also preferred ($p < .027$) over NONE for working on the same or shifting workspaces.

7 Discussion

This experiment examined the effect of SAR precue techniques for a procedural task on three different workspaces around the user. Many interesting observations can be made from the results of this study.

7.1 Comparing Multi-Level Precues: Same Workspace

Firstly, when observing the number of precues displayed simultaneously, differences were noted when working in the same or shifting workspace. When working in the same workspace (H1), additional cues predicting buttons to press beyond the target and upcoming buttons further enhanced response times for the TRN conditions. However, Multi-level 2 and 3 for SOL LINE saw a significantly slower reaction time than just displaying a single colored line. The decay in performance for LINE SOL is interesting as response times improve for their LINE TRN counterparts, implying LINE requires varying levels of transparency to be beneficial. The ARC conditions saw a similar effect. However, ARC SOL 2 was still faster than the single arc condition. Regarding errors, all cues showed a slight increase as the Multi-level factor increased. The largest increase was noticed in the NONE level 3



(a) Participant preference for precue and color combination, 1 best and 5 worst.

(b) Participant preference for the Multi-level precues, 1 best and 3 worst.

(c) Participant preference for precue style, 1 best and 3 worst. Working on the same and shifting workspace are combined.

Fig. 11: Post-study questions: participant preferences. Bars represent standard errors.

condition. For both response times and errors, 2 levels of precue was more beneficial compared to 3. Comparisons between the number of displayed precues indicate that presenting more precues can improve response times at a slight increase in errors, thus supporting H1 for various cues. However, increasing the level of precue may have a detrimental effect on performance, thus we suggest 2 levels of precue. Furthermore, there should be multiple factors to differentiate the correct order of precues displayed, in this case, size and color.

7.2 Comparing Multi-Level Precues: Shifting Workspace

When shifting to another workspace, increasing the number of precues significantly increased response times for the ARC and NONE conditions. We assume this is due to the additional precue information causing an extended search time for the initial button to press on the new workspace. However, LINE did not see a significant increase in response times, indicating it was easier to find the initial button to press. H2 is supported by ARC and NONE having significantly worse performance times as the number of precues increased. However, H2 is not supported by LINE as the LINE conditions did not decay in performance as heavily with additional precue levels. This finding also supports H4, as LINE was the only cue type that did not see a significant increase in response times when shifting workspaces.

7.3 Combining an ARC or LINE Precue with Color

When combining existing precue techniques, such as those from Volmer et al. [44], performance can be further improved. While existing research separated colored or directional-based cues, this work combined them. Specifically, the LINE and ARC conditions were merged with a singular color for highlighting the active task and multiple colors for upcoming tasks. LINE being combined with an active color or multiple transparencies of color saw different results. Compared to having just one line annotation, having just the active task with a different color decreased performance with multiple lines. Meanwhile, multiple color transparencies with several lines improved response time. Compared to prior research by Volmer et al. [44], ARC saw a benefit to response times by adding colored annotations. Response times are faster with the addition of single or multiple colored cues regarding response time, compared to prior studies. Furthermore, all LINE and ARC conditions saw an improvement to their NONE counterparts, with the exception of ARC SOL 3. Additionally, NONE generally saw a higher number of errors compared to ARC and LINE with the same Multi-level factor. The NONE condition was also significantly less preferred over conditions with a directional cue (Fig. 11a) and caused a higher amount of self-assessed mental effort (Fig. 10), thus supporting H3.

7.4 Conditions with Line Annotations

Further observations comparing the cue conditions show that LINE was superior regarding response times for working in the same or shifting workspaces. LINE conditions also saw few errors overall, and some required significantly less self-assessed mental effort than the NONE cue conditions. Post-questionnaire results also saw a strong preference towards the LINE conditions. The overall strong performance of the LINE conditions aligns with prior research suggesting drawing a line between the active task and upcoming task to be the most beneficial form of precue, validating H4.

7.5 Multiple Precues with Different Color Transparencies

While LINE cues were overall superior, ARC TRN 2 was not significantly slower than the LINE SOL conditions when working in the same workspace. This is an interesting observation as it implies adding multiple color transparencies to ARC can improve response times to the performance of LINE with one active color. LINE and ARC TRN conditions were also significantly faster than their SOL counterparts when working in the same workspace. This further demonstrates that having progressively more transparent shades of color for all cues improves response times compared to just changing the color of the active task. While there is a slight increase in errors for the TRN conditions for ARC and LINE, there is still an overall low number of errors. Additionally, when shifting to another workspace, response times for TRN became slightly slower than their SOL equivalents. However, when analyzing the same Multi-level factor against each other, no SOL to TRN comparisons were deemed significant for both ARC and LINE cues. Therefore, we believe H5 is supported by the significantly faster performance of the TRN conditions when working in the same workspace. Although, H5 is not supported for shifting workspaces as there was no significant difference.

7.6 Comparing Multi-Level Precues: Participant Preference

Fig. 11 depicts the results of the post-study questionnaire. Participants indicated LINE with SOL and LINE with TRN were their preferred combinations, see Fig. 11a. Observations from participant preference for the number of Multi-level precues (see Fig. 11b) only revealed significant differences for NONE. Participants preferred 1 level of cue compared to having 3. To a lesser extent, participants also preferred having 2 colored transparent annotations over 3. The ARC and LINE conditions saw no significant difference between levels of precue. This indicates including a directional annotation improves user preference when utilizing multiple levels of precues. As depicted in Fig. 11c, LINE was the preferred precue.

7.7 Highlighting the Position of the Upcoming Workspace

We hypothesized (H6) that cueing the position in participants' peripheral vision could draw their attention, similar to Bailey et al. [3], and they would move their hand and head more efficiently. However, the participants moved both their hand and head faster in LINE than in ARC and NONE (see Fig. 9). Therefore, H6 is not supported. This might mean that showing the place of the upcoming workspace can draw their attention, but showing the path to the next workspace will be more beneficial for the participants. The participants moved the second fastest in ARC and slowest in NONE. This indicates *path* information provided the most help for participants, followed by *direction* information, and then *place* information. This trend is similar to the finding by Liu et al. [25].

7.8 Comparing Different Physical Workspace Devices

Different button device panels were utilized in this study. We made observations comparing the box and dome devices. Supplementary material Figures A–D showcase the differences in response time and errors between conditions and workspaces. For response times, all conditions had similar performance when working in the same workspace or switching between workspaces, except for the NONE conditions, which were slower when switching to the box (Figure B in supplementary material). This may have been caused by highlights in the periphery being easier to see on the curved dome surface than on the flat box, resulting in the participant taking more time to switch.

Comparing errors (Tab. 4), the box saw a lower percentage of errors compared to both dome stations when working on the same device or shifting to another one (Sec. 6.3). While few errors were made across all devices, we believe the small jump in errors for the dome devices is due to their curved surface. As the box is a flat surface, the user is able to view all buttons and visualizations evenly from looking above. Meanwhile, for both domes, it is harder for the user to clearly see the annotations that were not in their center field of view. However, we are not able to find different user behaviors on each device using the head and hand tracking data mentioned in Sec. 5.3. This may be because the differences were reflected in eye movement rather than head or hand movement. Future work could use eye tracking to compare how the participant's gaze trace is affected by different device shapes.

7.9 Limitations

While this study presents precues in multiple workspaces around the user, there are limitations to what we explore. First, the configuration of the devices in the workspace has an impact on the effectiveness of the precues. Examining larger and more complex workspaces is a future research direction.

Next, the devices utilized in this experiment were painted white and optimized for viewing the projected annotations. Surfaces may have varying levels of occlusion and reflectivity, which would impact the SAR precue annotations. Future work could investigate precues on more complex surfaces.

Finally, the tasks performed in this experiment were simple button presses completed within seconds. The effectiveness of precues needs to be investigated in more complex tasks, such as those involved in equipment assembly, maintenance, and repair [18].

7.10 Summary

Overall, this study investigated numerous aspects behind SAR precues on multiple workspaces. Akin to existing research, having any form of precue was better than just highlighting the active task when working in the same workspace. Furthermore, linking lines between the current and upcoming tasks was considered superior for working in the same or shifting to another workspace. Providing multiple levels of color transparency was also discovered to further improve the response times of the directional-precues.

8 Conclusions and Future Work

We investigated SAR precues to guide users throughout multiple workspaces. While existing work utilized direction-based and target-based cues to guide to the upcoming task [43–45], we combined the benefits of both cueing styles into one and evaluated them using a procedural task. This work also took inspiration from VR precueing research that indicated displaying precues multiple steps in advance increased performance [25].

This study confirmed that precueing multiple steps in advance can enhance performance for SAR procedural tasks in the same workspace. These precues should be more transparent than the active task and the same color. While multiple precues can benefit performance in the same workspace, this additional information can make shifting workspaces more difficult. The LINE conditions did not appear to be significantly affected by this regarding response times. Akin to previous SAR precue studies, conditions with a LINE were found to be the most beneficial for tasks within the same workspace. LINE is also superior when shifting to another workspace.

Overall, to enhance user performance for procedural tasks in the same workspace, our results suggest utilizing two or three directional LINE cues of different widths and shades of the same color. In addition, when shifting workspaces, using one, two, or three LINE annotations can improve response times. However, only the active task should be a different color to improve accuracy, as it is easier to detect when working at a new location.

Future research could examine different precue visualizations and investigate them in other environments. While the NONE workspace-shift cue highlighted the entire upcoming workspace was not beneficial compared to the directional annotations, users could still notice it. Variations such as combining the workspace-shift cues of LINE and ARC with highlighting the upcoming workspace may produce different results.

Investigating SAR precues on different device surfaces may also reveal interesting observations. In this research, we conducted a brief analysis of two types of devices. However, a more in-depth study with a wider range of devices may uncover behavioral differences for the precues. Finally, utilizing the cues in a single large workspace with tasks outside the near field of view may have different observations than several small workspaces.

Acknowledgments

Volmer and Thomas were partially funded by the Australian Research Council Discovery Grant DP180100755 and the Australian Research Centre for Interactive and Virtual Environments at the University of South Australia. Liu and Feiner were funded in part by National Science Foundation Grant CMMI-2037101.

References

[1] D. Andersen, V. Popescu, M. E. Cabrera, A. Shanghavi, G. Gomez, S. Marley, B. Mullis, and J. Wachs. Virtual annotations of the surgical field through an augmented reality transparent display. *The Visual Computer*, 32:1481–1498, 2016. doi: 10.1007/s00371-015-1135-6 1

[2] R. Azuma and C. Furmanski. Evaluating label placement for augmented reality view management. In *The Second IEEE and ACM International Symposium on Mixed and Augmented Reality, 2003.*, pp. 66–75. IEEE, 2003. doi: 10.1109/ISMAR.2003.1240689 1

[3] R. Bailey, A. McNamara, N. Sudarsanam, and C. Grimm. Subtle gaze direction. *ACM Transactions on Graphics*, 28(4), sep 2009. doi: 10.1145/1559755.1559757 4, 8

[4] P. Baudisch and R. Rosenholtz. Halo: A technique for visualizing off-screen objects. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, pp. 481–488. Association for Computing Machinery, New York, NY, USA, 2003. doi: 10.1145/642611.642695 2

[5] J. Baumeister, J. Dorrian, S. Banks, A. Chatburn, R. T. Smith, M. A. Carskadon, K. Lushington, and B. H. Thomas. Augmented reality as a countermeasure for sleep deprivation. *IEEE Transactions on Visualization and Computer Graphics*, 22(4):1396–1405, 2016. doi: 10.1109/TVCG.2016.2518133 1, 2

[6] J. Baumeister, S. Y. Ssin, N. A. M. ElSayed, J. Dorrian, D. P. Webb, J. A. Walsh, T. M. Simon, A. Irlitti, R. T. Smith, M. Kohler, and B. H. Thomas. Cognitive cost of using augmented reality displays. *IEEE Transactions on Visualization and Computer Graphics*, 23(11):2378–2388, 2017. doi: 10.1109/TVCG.2017.2735098 2

[7] M. Billinghurst, A. Clark, and G. Lee. A survey of augmented reality. *Foundations and Trends in Human–Computer Interaction*, 8(2-3):73–272, 2015. doi: 10.1561/1100000049 1

[8] O. Bimber and R. Raskar. *Spatial Augmented Reality: Merging Real and Virtual Worlds*. AK Peters/CRC Press, 2005. doi: 10.1201/b10624 1

[9] J. Carmigniani, B. Furht, M. Anisetti, P. Ceravolo, E. Damiani, and M. Ivkovic. Augmented reality technologies, systems and applications. *Multimedia Tools and Applications*, 51:341–377, 2011. doi: 10.1007/s11042-010-0660-6 1

[10] T. P. Caudell and D. W. Mizell. Augmented reality: an application of heads-up display technology to manual manufacturing processes. In *Proceedings*

of the Twenty-Fifth Hawaii International Conference on System Sciences, vol. ii, pp. 659–669 vol.2, 1992. doi: 10.1109/HICSS.1992.183317 1

[11] S. Condino, M. Carbone, R. Piazza, M. Ferrari, and V. Ferrari. Perceptual limits of optical see-through visors for augmented reality guidance of manual tasks. *IEEE Transactions on Biomedical Engineering*, 67(2):411–419, 2020. doi: 10.1109/TBME.2019.2914517 2

[12] A. Doshi, R. T. Smith, B. H. Thomas, and C. Bouras. Use of projector based augmented reality to improve manual spot-welding precision and accuracy for automotive manufacturing. *The International Journal of Advanced Manufacturing Technology*, 89:1279–1293, 2017. doi: 10.1007/s00170-016-9164-5 1

[13] E. Eiriksdottir and R. Catrambone. Procedural instructions, principles, and examples: How to structure instructions for procedural tasks to enhance performance, learning, and transfer. *Human factors*, 53(6):749–770, 2011. doi: 10.1177/0018720811419154 1

[14] S. Feiner, B. Macintyre, and D. Seligmann. Knowledge-based augmented reality. *Communications of the ACM*, 36(7):53–62, jul 1993. doi: 10.1145/159544.159587 1

[15] B. Furht. *Handbook of Augmented Reality*. Springer Science & Business Media, 2011. doi: 10.1007/978-1-4614-0064-6 1

[16] F. A. Hansen. Ubiquitous annotation systems: technologies and challenges. In *Proceedings of the Seventeenth Conference on Hypertext and Hypermedia*, pp. 121–132. ACM, 2006. doi: 10.1145/1149941.1149967 1, 2

[17] F. Heinrich, F. Joeres, K. Lawonn, and C. Hansen. Comparison of projective augmented reality concepts to support medical needle insertion. *IEEE Transactions on Visualization and Computer Graphics*, 25(6):2157–2167, 2019. doi: 10.1109/TVCG.2019.2903942 2

[18] S. J. Henderson and S. K. Feiner. Augmented reality in the psychomotor phase of a procedural task. In *2011 10th IEEE International Symposium on Mixed and Augmented Reality*, pp. 191–200. IEEE, 2011. doi: 10.1109/ISMAR.2011.6092386 9

[19] 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 1

[20] R. Hoffmann, P. Baudisch, and D. S. Weld. Evaluating visual cues for window switching on large screens. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '08, p. 929–938. Association for Computing Machinery, New York, NY, USA, 2008. doi: 10.1145/1357054.1357199 2

[21] F. Jasche, S. Hoffmann, T. Ludwig, and V. Wulf. Comparison of different types of augmented reality visualizations for instructions. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3411764.3445724 1

[22] R. Krempien, H. Hoppe, L. Kahrs, S. Daeuber, O. Schorr, G. Eggers, M. Bischof, M. W. Munter, J. Debus, and W. Harms. Projector-based augmented reality for intuitive intraoperative guidance in image-guided 3D interstitial brachytherapy. *International Journal of Radiation Oncology*Biology*Physics*, 70(3):944–952, 2008. doi: 10.1016/j.ijrobp.2007.10.048 2

[23] H. Lee and W. C. Cha. Virtual reality-based ergonomic modeling and evaluation framework for nuclear power plant operation and control. *Sustainability*, 11(9), 2019. doi: 10.3390/su11092630 1

[24] 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, pp. 2973—-2976. Association for Computing Machinery, New York, NY, USA, 2012. doi: 10.1145/2207676.2208706 1

[25] 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 1, 2, 3, 4, 9

[26] 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 1, 2

[27] M. R. Marner, A. Irlitti, and B. H. Thomas. Improving procedural task performance with augmented reality annotations. In *IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 39–48. IEEE, 2013. doi: 10.1109/ISMAR.2013.6671762 1, 2

[28] MathWorks. *Matlab*. `https://www.mathworks.com/`, last accessed on 03/25/2023. 5

[29] A. Mewes, F. Heinrich, U. Kägebein, B. Hensen, F. Wacker, and C. Hansen. Projector-based augmented reality system for interventional visualization inside MRI scanners. *The International Journal of Medical Robotics and Computer Assisted Surgery*, 15(1):e1950, 2019. doi: 10.1002/rcs.1950 1, 2

[30] NaturalPoint Inc. *OptiTrack Flex 13*. `https://optitrack.com/cameras/flex-13/`, last accessed on 03/25/2023. 4

[31] F. G. Paas. Training strategies for attaining transfer of problem-solving skill in statistics: A cognitive-load approach. *Journal of Educational Psychology*, 84(4):429, 1992. doi: 10.1037/0022-0663.84.4.429 4, 7

[32] P. Parekh, S. Patel, N. Patel, and M. Shah. Systematic review and meta-analysis of augmented reality in medicine, retail, and games. *Visual Computing for Industry, Biomedicine, and Art*, 3:1–20, 2020. doi: 10.1186/s42492-020-00057-7 1

[33] H. Pashler. Dual-task interference in simple tasks: data and theory. *Psychological Bulletin*, 116(2):220, 1994. doi: 10.1037/0033-2909.116.2.220 2

[34] K. Pfeil, S. Masnadi, J. Belga, J.-V. T. Sera-Josef, and J. LaViola. Distance perception with a video see-through head-mounted display. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, CHI '21. Association for Computing Machinery, New York, NY, USA, 2021. doi: 10.1145/3411764.3445223 2

[35] 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, 2017. doi: 10.1109/TVCG.2017.2709746 1

[36] F. Putze, A. Vourvopoulos, A. Lécuyer, D. Krusienski, S. Bermúdez i Badia, T. Mullen, and C. Herff. Editorial: Brain-computer interfaces and augmented/virtual reality. *Frontiers in Human Neuroscience*, 14, 2020. doi: 10.3389/fnhum.2020.00144 2

[37] R. Raskar, G. Welch, and H. Fuchs. Spatially augmented reality. In *Proceedings of the International Workshop on Augmented Reality: Placing Artificial Objects in Real Scenes: Placing Artificial Objects in Real Scenes*, IWAR '98, p. 63–72. A. K. Peters, Ltd., USA, 1999. 1

[38] J. Rekimoto and M. Saitoh. Augmented surfaces: A spatially continuous work space for hybrid computing environments. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '99, p. 378–385. Association for Computing Machinery, New York, NY, USA, 1999. doi: 10.1145/302979.303113 2

[39] J. Serván, F. Mas, J. L. Menéndez, and J. Ríos. Using augmented reality in Airbus A400M shop floor assembly work instructions. *AIP Conference Proceedings*, 1431(1):633–640, 2012. doi: 10.1063/1.4707618 1

[40] A. Tang, C. Owen, F. Biocca, and W. Mou. Comparative effectiveness of augmented reality in object assembly. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*, CHI '03, pp. 73––80. Association for Computing Machinery, New York, NY, USA, 2003. doi: 10.1145/642611.642626 1

[41] J. W. Tukey. Comparing individual means in the analysis of variance. *Biometrics*, 5(2):99–114, 1949. doi: 10.2307/3001913 5

[42] Unity. *Unity*. `https://unity.com/`, last accessed on 03/25/2023. 4

[43] B. Volmer, J. Baumeister, R. Matthews, L. Grosser, S. Von Itzstein, S. Banks, and B. H. Thomas. A comparison of spatial augmented reality predictive cues and their effects on sleep deprived users. In *IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 589–598. IEEE, 2022. doi: 10.1109/VR51125.2022.00079 1, 2, 9

[44] 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 1, 2, 3, 4, 8, 9

[45] 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*, 2022. doi: 10.1109/TVCG.2022.3197810 2, 9

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

[47] L.-C. Wu, I.-C. Lin, and M.-H. Tsai. Augmented reality instruction for object assembly based on markerless tracking. In *Proceedings of the 20th ACM SIGGRAPH Symposium on Interactive 3D Graphics and Games*, I3D '16, pp. 95–102. Association for Computing Machinery, New York,

NY, USA, 2016. doi: 10.1145/2856400.2856416 1

[48] C.-s. Yu, E. M.-y. Wang, W.-C. Li, and G. Braithwaite. Pilots visual scan patterns and situation awareness in flight operations. *Aviation, Space, and Environmental Medicine*, 85(7):708–714, 2014. doi: doi:10.3357/ASEM.3847.2014 1

[49] 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 1

[50] F. Zhou, H. B.-L. Duh, and M. Billinghurst. Trends in augmented reality tracking, interaction and display: A review of ten years of ISMAR. In *2008 7th IEEE/ACM International Symposium on Mixed and Augmented Reality*, pp. 193–202, 2008. doi: 10.1109/ISMAR.2008.4637362 1