Contents lists available at ScienceDirect

Computers & Graphics

journal homepage: www.elsevier.com/locate/cag

Visual Cues in VR for Guiding Attention vs. Restoring Attention after a Short Distraction

Jason W. Woodworth[a,*], Christoph W. Borst[a,1]

[a] CACS VR Lab, University of Louisiana at Lafayette, Louisiana, USA

ARTICLE INFO

Article history:
Received September 5, 2025

Keywords: Virtual Reality, Attention Guidance, Attention Restoration, Visual Cues

This is an author-formatted version. Original publication: J. W. Woodworth and C. W. Borst, "Visual Cues in VR for Guiding Attention vs. Restoring Attention after a Short Distraction" Computers & Graphics 118, 2024
doi: 10.1016/j.cag.2023.12.008

ABSTRACT

Distraction in VR training environments may be mitigated with a visual cue intended to guide user attention to a target. A survey of related literature suggests a past focus on "search and selection" tasks to evaluate a cue's capability for guidance. We investigate the capability of 9 eye-tracked cues with a new type of task that focuses on how to restore attention when a short distraction (e.g., a notification) shifts focus away from a target. Our study includes a guidance task in which subjects gaze at objects in a randomized order and a restoration task in which gaze sequences are interrupted by distraction events after which gaze must be returned to an object. We consider a wider variety of factors and metrics than previous studies, varying object spacing, gaze dwell time, and distraction distance and duration, and breaking down guidance time into subcomponents. Results show a general positive trend for cues that directly connect the user's gaze to the target rather than indirectly suggesting direction. Results further reveal different patterns of cue effectiveness for the restoration task than for conventional guidance. This may be attributed to knowledge that subjects have about the location of the object from which they were distracted. An implication for more complex distraction tasks is that we expect them to be between the short distraction and regular guidance in terms of memory of object position. So, we speculate cue performance for other tasks would vary between the short distraction and guidance results. For restoration, some cues add complexity that reduces, rather than improves, performance.

1. Introduction

As virtual and augmented reality technologies have matured, they have seen a rapid increase in application to industry [1, 2]. VR can be used to simulate dangerous situations with minimal risk for training purposes [3] or improve designs off-site [4], while AR can be used to augment the work site with helpful guidance imagery [5, 6]. In each scenario, the user's visual attention is of paramount importance [7]. For example, a trainee in a VR oil rig training simulation may be distracted by system alerts and miss important details while not looking at a target object [3], a repair technician may not be able to find the target area of interest [5], and a warehouse employee may need guidance to find the next part in an order picking task [8]. Visual cues, such as arrows [9, 10, 11] or highlights [12, 13] on target objects, are often employed in such systems to guide a user's attention to relevant objects, and as such have been the subject of substantial research.

Attention restoration after a distraction, however, presents a more dangerous scenario in the industrial context. A construction worker, for example, can be injured after a short loss of situational awareness on-site [7]. This could be mitigated by visual cues restoring attention to an important target after a distraction. However, such cues have primarily been studied in a guidance context in which the cue guides the user to a new, previously unseen target, while less work has focused on how to

*Corresponding author:
e-mail: jason.woodworth1@louisiana.edu (Jason Woodworth)
[1] e-mail: cwborst@gmail.com

guide a user's attention back to a target once it has already been lost.

We consider that these two contexts may require different cue solutions, and so to address this gap, we present user studies that evaluate the performance of 9 visual cues at restoring visual attention to a target object when their attention shifts elsewhere (attention restoration) and simply guide users' attention to another object (attention guidance). Our results give insight into the differences between these tasks and why results may not be directly transferable from one to the other.

Cues were first tuned to subjectively preferred parameters during a preliminary tuning study (see Appendix A). In a separate main study, attention guidance was tested by cues visually guiding users to a set of targets in a random sequence. During this we measured total time to get through a whole set and total number of incorrect targets looked at during the trial. To test attention restoration, this sequence was occasionally interrupted by distraction events, after which gaze was restored by a visual cue. During this we measured the time taken to look back at the correct target. To better understand the nature of the tasks and the cues' performance in different scenarios, we also varied task parameters such as target spacing (wide or narrow), amount of gaze-dwell time required to activate a target (short or long), distance that distractions spawn from the target (close or far), and number of distraction elements (1, 3, or 5).

Through our study results, we find that a highlight, DynSWAVE [14], and a thick line strip rendered along a cubic curve are generally the best performing cues, though each performed best in different scenarios. We also find a different pattern of cue effectiveness between restoration and guidance tasks, suggesting a fundamental difference in how solutions for them should be approached.

We consider the contributions of our work to be: 1) An extensive direct comparison of recent prior studies on attention guidance and restoration cues, 2) a comparison of the largest collection of cues available in the literature, 3) insight into the difference between restoration and guidance tasks, and 4) deeper insight into how each cue performs in varying conditions (wide vs narrow target layouts, different distraction parameters) and recommendations for particular training contexts.

This paper extends work that was presented in [15]. We now include a larger study analysis that breaks down guidance time into subcomponents, considers errors made during guidance and restoration stages, and reports subjective feedback. Additionally, we include results from the preliminary study in which users tuned parameters of each cue.

2. Related Works

Many approaches have been suggested for guiding a user's attention to nearby or out-of-view targets. In Table 1, we present a comparison of the most relevant works in the literature. We prioritize recent works and those that compare a large number of cues. While our work focuses on VR, we include works in other mediums (e.g. augmented reality) because they solve a similar problem. We present observations and patterns found from looking at the state of the field as a whole.

A form of arrow pointing at the target is by far the most commonly tested visual cue, appearing in over half (12) of the surveyed works, and is often used as a baseline to test a new cue against (e.g. [10, 19, 27]). Arrows typically perform well [27, 14], but not exceptionally so, and may have difficulty guiding to targets behind the user [23]. Arrows are most often rendered as 3D objects pointing at the object's position in the scene [10, 25] or a 2D image rendered pointing in the direction the user should look to face the target [26, 14]. The FlyingARrow [11, 24], a more novel arrow technique, animates the arrow to move from the user's view towards the target, but the cue's animation speed may affect the user's eye movement speed.

A form of target highlighting is also among the most common guidance techniques. Classically highlights take the form of a simple visual overlay or outline of the target object [29, 30, 31]. More subtle highlighting cues often leverage the human visual periphery's sensitivity to flickering to draw gaze to the target without the user even noticing the highlight was there [16, 32]. Other works insert highlights diegetically (e.g. a firefly [13] or swarm of bugs [12]) to draw attention while maintaining immersion.

A common limitation with highlighting techniques is the requirement for the target to already be in the user's field of view. If the user is looking far enough away from the target and cannot see the cue, it will have no effect. Grogorick et al. [32] attempted to remedy this by moving the flickering stimulus towards the target, but study on this technique has been limited.

The most common task to compare and evaluate these visual cues is Search and Selection. In most, a target is selected by the system, the cue guides the subject's gaze towards it, and the subject must indicate that they found it. Distractors are often present to make the task more challenging. Targets are often spread 360° around the subject [20, 18, 10], but are sometimes just placed in front [19, 12, 27] if the author's context deems it appropriate.

We consider these tasks to be centered on attention guidance; critically, the subject does not have previous knowledge of the target's position before the cue appears. We argue that this is fundamentally different from attention restoration, in which the user has already looked at the target but was visually distracted and must be guided back. In our survey, we found no works directly addressing the problem of attention restoration, which we address in our following study.

Despite the increased accessibility of eye-tracking within headsets, most (14) of the surveyed papers did not include eye-tracking within their cues' designs. That is, the appearance of the cues did not respond to the user's eye movements, and was instead placed statically in a scene or fixed to head gaze. Studies that included eye-tracking suggest that designing cues to respond to eye movements has potential to improve visual cues [14]; therefore we include eye-tracking in the design of all of our compared cues.

Our study addresses the noted gaps in the literature by introducing a new type of task that mirrors minor distractions (e.g. phone notifications) and by directly comparing the largest selection of cues in the literature (9 plus a No Cue baseline), under a wide variety of conditions (e.g. different target layouts).

Author	Cues	Task	Metrics	Results	ET	Subjects	Medium
Bailey, 2009 [16]	SGD (warm-cool), SGD (luminance), None	Image viewing; SGD occasionally appears in unusual areas	Response time, perceived image quality, gaze patterns	SGD produced low response time (<500ms) and effectively changed gaze patterns	Yes	10	Desktop
Binetti, 2021 [9]	**Arrow** (near and far depths), Auditory	Search and selection; targets spread 360° around subject; accuracy not enforced	Selection time and accuracy, NASA-TLX	Near arrow showed worse times, but was alleviated with added auditory cue	No	32	AR
Biocca, 2006 [17]	**Attention Funnel**, **Highlight**, Verbal Description	Search and selection; targets spread 360° around subject; accuracy not enforced	Selection time, error, and variability, NASA-TLX	Attention Funnel enabled faster target selection, higher consistency, and demanded least mental effort	No	14	AR
Bork, 2018 [10]	**3D Arrow**, Aroundplot, 3D Radar, EyeSee360, sidebARs, Mirror Ball	Search and selection with varying target patterns; targets spread 360° around subject; accuracy enforced	Selection time, Head trajectory pattern	EyeSee360 and 3D Radar showed lowest selection times	No	24	AR
Grogorick, 2018 [18]	ColorDot, SGD, ZoomRect, ZoomCircle, SpatialBlur, None	360 ° image viewing, cues occasionally appear to attract gaze	Search time, Subtlety of cue	No cues performed outstandingly well; ZoomRect performed best	Yes	102	VR
Gruenefeld, 2017 [19]	**Arrow**, **Halo**, Wedge, EyeSee360	Direction estimation; targets 180° or 360° around subject; accuracy not enforced	Estimation accuracy, NASA-TLX	EyeSee360 showed lowest estimation error; No result for task load	No	16	AR
Gruenefeld, 2018 [20]	**Halo** (AR and VR), Wedge (AR and VR)	Search and selection; Direction estimation; Targets spread 360° around subject; accuracy not enforced	Selection time and accuracy, Estimation error, SUS, RAW-TLX	Halo and Wedge performed similarly well; Correlation found between number of targets and selection time	No	16	VR & AR
Gruenefeld, 2018 [11]	FlyingARrow, EyeSee360	Search and selection; Direction estimation; Targets spread 360° around subject; accuracy not enforced	Selection time and error, Estimation error, SUS, RAW-TLX	FlyingARrow showed slightly worse selection time and estimation, but not significantly so, but was rated more usable	No	12	AR
Gruenefeld, 2019 [21]	3D Radar, EyeSee360	Movement estimation of out-of-view targets	Estimation error, usability	3D Radar enabled more accurate estimation and was considered more usable	No	48	VR
Gugenheimer, 2016 [22]	SwiVRChair, None	360° video viewing with story stimuli surrounding subject; occasional forced rotation and blocks	RSSQ, Presence, Enjoyment	SwiVRChair produced higher presence and enjoyment scores; Both resulted in minimal simulator sickness	No	16	VR
Harada, 2022 [23]	Moving Window, **3D Arrow**, Radiation, Spherical Color Gradiation, 3D Radar	Search task; Targets spread 360° around subject; accuracy not enforced	Search time, Cue recognition time, Fixation count and duration, Saccade count and length	Moving Window and Radiation had low search time for frontal targets but high times for targets in back; 3D Radar showed even search times across all regions	No	30	VR
Hu, 2021 [24]	bSOUS, fSOUS, FlyingARrow (+arc and +trail)	Search and selection; Targets generally in front of subject; accuracy enforced, but targets disappeared after time	Search time, Selection success rate, NASA-TLX	bSOUS generally outperformed fSOUS and incurred lower task load; FlyingARrow performed second-best, but cue speed may influence search time	No	24	VR
Jo, 2011 [25]	Aroundplot, 2D Radar, 3D Arrow Cluster	Search task; targets spread 360° around subject; accuracy enforced in 1 of 2 tasks	Search time, Search failure rate, NASA-TLX	Aroundplot enabled a lower search time and fail rate when more objects were present compared to 2D Radar	No	16	Mobile AR
Lange, 2020 [12]	HiveFive, **Arrow**, Blurring, Deadeye, SGD, None	Search and selection; Targets generally in front of subject; accuracy not enforced	Time to First Fixation (TTFF), Selection accuracy, IPQ, RAW-TLX	HiveFive showed lowest TTFF and highest immersion scores; Arrow showed similarly low TTFF	No	20	VR
Lin, 2017 [26]	AutoPilot, **Arrow**, None	Viewing 360° videos with different desired gaze patterns	Ease of focus, Engagement, Enjoyment, Presence, Discomfort	Cues increased enjoyment and ease of focus; AutoPilot resulted in discomfort with large rotations	No	32	VR
Markov-Vetter, 2020 [27]	MAP, MAP+, **Arrow**	Search and selection; Targets in a ring in front of subject; Secondary oddball task included for task load	Search speed, False selections, Errors in oddball task, HRV, NASA-TLX	Arrow enabled faster search speeds and lower task load. Objective metrics of task load were more sensitive.	No	15	AR
Renner, 2017 [14]	None, **Attention Funnel**, Arrow, Arrow_ET Flickering, SWave, **DynSWave**	Search and selection task; Targets to subject's left and right; accuracy enforced	Selection time, Angular head movement	Arrow was fastest and best rated, with SWave showing similar results; DynSWave performed worse than SWave	Yes	20	VAR
Renner, 2017 [28]	Picking Light, Arrow, SWave	Guided placement; Cues placed either in center of FOV or upper right	Placement time, Head movements, NASA-TLX	Cues in central FOV performed better than cues in periphery; SWave outperformed arrow	Yes	21	VAR

Table 1. Works related to the field of attention guidance cues, including a brief description of the associated task and main results, and whether Eye Tracking (ET) was used in cue design. Works comparing multiple cues in mixed reality were prioritized. Bolding indicates the cue, or a variation of it, was used in our study. Medium indicates how the cues were visualized, with AR implying use of an AR headset while mobile AR implies a tablet or phone.

3. Visual Cues Design

Figure 1 shows visual cues in the test environment with Narrow Target Spacing. Cues were tuned to reasonable appearance parameters within a preliminary study described in Appendix A. Cues are:

No Cue: We consider the absence of a cue as a baseline.

Arrow (Arr): The *Arrow* (Figure 1a) is a single 3D arrow. Such arrows are perhaps the most commonly deployed cue (e.g., Google Expeditions). The arrow is placed along the path of the shortest arc on a head-centered sphere with sphere radius being the head-to-target distance.

Arrow Trail (ArrT): The *Arrow Trail* (Figure 1b) places multiple arrows along the same arc as arrow, forming a trail from the gaze point to the target. As the gaze moves away from the target, the number of arrows increases. This may provide a stronger cue than a single arrow.

Arrow Field (ArrF): The *Arrow Field* (Figure 1c) is a novel cue that renders many arrows on a head-centered sphere. Arrow placement does not depend on any gaze tracking: arrows appear static rather than moving. The arrows appear at vertices of an (invisible) icosphere and each points along an arc towards the target. This may be useful with reduced or no eye tracking.

DynSWAVE (DynS): *DynSWAVE* (Figure 1d) is based on a cue by Renner and Pfeiffer [14]. Concentric circles move along a head-centered sphere towards the target, such that the movement appears like decreasing circle radius (implemented by scrolling textured lines towards a sphere pole anchored at the target). Movement speed increases as the user looks farther from the target. This cue performed well in work by others [14, 28], motivating its comparison to a broader set of cues.

Border (Bor): The *Border* (Figure 1e) displays a single circle on the head-centered sphere, so the circle appears centered on the target. As gaze moves away from the target, the circle's edge follows the gaze (radius appears to increase) until it reaches a boundary. This cue is similar to Halo [20], however, considering our experiment environment, we do not adjust for clutter.

Vignette (Vig): Our *Vignette* (Figure 1f) cue is similar to Fade to Black [33]; it progressively darkens portions of the scene far from the target as user attention drifts. We use a screen-space shader to orient a tunnel effect in front of the user (with an offset toward the target), which gets more prominent (darker, sharper) the farther the user looks away. The open portion of the tunnel never leaves the user's field of view completely, to prevent confusion about where to shift visual focus.

Line Strip (LineS): The *Line Strip* (Figure 1g) is a simple cue sharing some features with Wedge [20] and 3DPath [34]. A thick line strip (appearing like a curved banner) is displayed along a 3D Hermite cubic curve from a point on the gaze vector (at a small distance from the head) to the target, as in Biocca et al. [17], showing the full path. The simple view of the whole path may be effective or liked for its clarity.

Attention Funnel (AtFun): The *Attention Funnel* (Figure 1h) is based on Biocca et al. [17], which placed multiple rings along a cubic curve between the student and the target.

Highlight (Hi): *Highlight* (Figure 1i) is a simple cue that highlights the target with a yellow emission effect, increasing in intensity as the eye moves away from the target. The color was chosen to contrast other scene colors. Unlike the other cues, this cue appears directly on the target, which may draw attention immediately to the target location as long as it is in the field of view. The highlight cue was not included in Table A.8 or the tuning study, because it was added for the main comparison study.

Cues other than Highlight and Vignette were colored red to contrast other scene colors and because red is a standard alert color for attention. Cues were rendered in such a way to be visible even behind occluding objects (occluding objects appear partially transparent, as for menus in [35]). A first-order IIR filter was applied to eye gaze when rendering cues to reduce high-frequency jitter and smooth its motion.

4. Guidance and Restoration Studies

Our experiment's main contribution was to study performance with an attention restoration task, different from prior studies. A guidance task was also included to clarify how different tasks produce difference relative cue performance, and to consider more conditions than prior guidance studies. The tasks required subjects to look at barrel targets in sequence, dwelling on them for a short time to activate and proceed to the next step.

In a guidance stage, subjects activated 8 targets in randomized order with a cue for guidance. In an attention-restoration stage, subjects occasionally had to shift gaze to spherical distraction objects and then return to the same target (with a cue active). Both stages used within-subjects design for 10 cue conditions (the 9 cues and No Cue) with other factors varying target and distraction properties. Both stages took place in a virtual offshore oil rig, with users seated at a station near the top of the rig and blue barrels used as targets (Figure 2). In both stages, targets or distractions were considered "looked at" if the angle between gaze vector and vector to target went below a 4 degree threshold.

4.1. Participants and Apparatus

98 subjects were recruited from local science and engineering departments for guidance and restoration studies, with ages ranging from 18 to 35 (mean 21.4, StDev 3). Due to a recording error, data from the first 31 subjects was not available for the restoration stage analysis. All subjects used an HTC Vive Pro Eye headset with one HTC Vive controller (for answering embedded questions). Experiments were performed on an Alienware Aurora R13 with an Intel 17-12700KF processor and GeForce RTX 3080 graphics card. The study was approved by the university Institutional Review Board under the approval number SU18-03 CACS.

4.2. Barrel Setup and Patterns

Two barrel setups (Figure 2), with varying target spacing (narrow or wide) were used. The subject was seated on a platform, 15 meters from the center of the grid of barrels. The narrow setup had 8 barrels in a 3x3 grid format with a missing

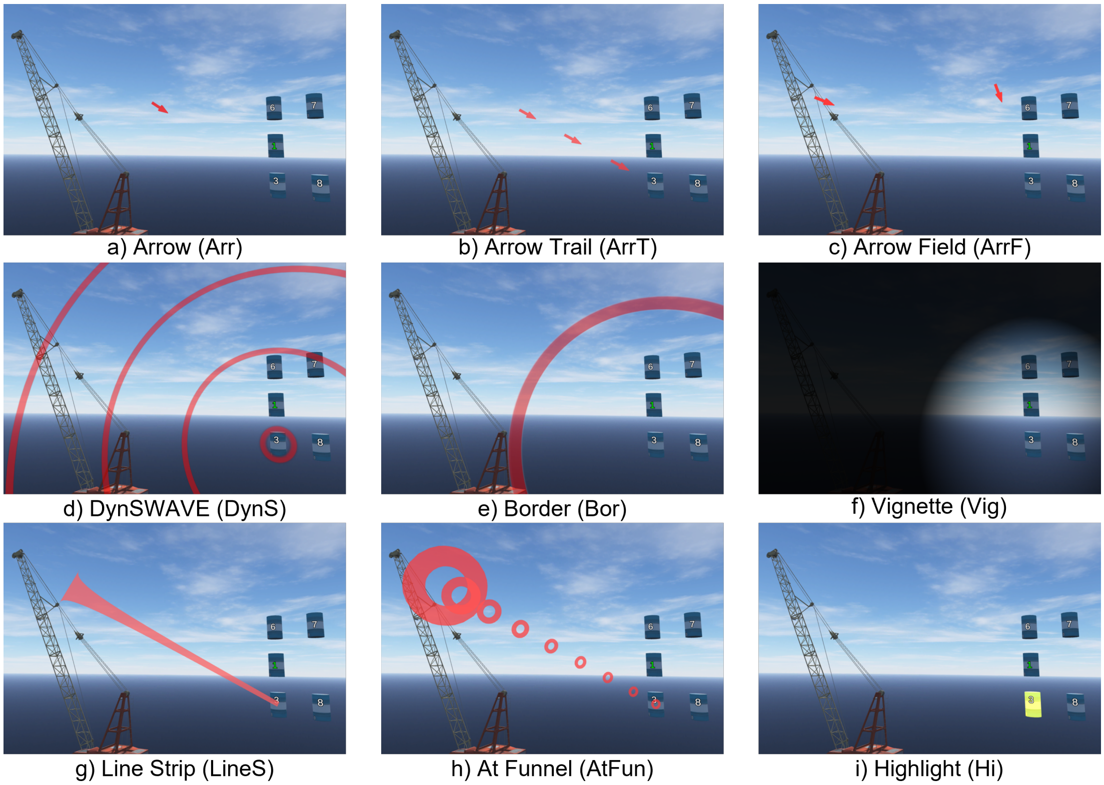

Fig. 1. The 9 visual cues compared in these studies. All examples are captured with the same gaze vector and target.

center, with each barrel 2 meters away from its neighboring barrels. The wide setup had 8 barrels in the same layout but spaced 6 meters away from their neighboring barrels.

Each barrel has a number between 1 and 8 displayed as a label in front. For 8 barrels, there are (8!) orders in which they can be sequenced. To ensure similar movement paths between barrel orders, we select randomly from a subset of 32 sequences that each consist of the follow movement types: 1 vertical move between neighboring barrels, 1 neighboring-barrel horizontal move, 1 long diagonal move (between non-neighbors), 1 two-unit vertical move (non-neighbors), 1 two-unit horizontal move (non-neighbors), 2 mixed moves crossing two horizontal and one vertical units, and total path length of 13.3 grid units. This made paths similar in complexity and length while appearing random to users.

4.3. Procedure

After filling out a consent form, a background questionnaire, and undergoing eye tracking calibration, subjects performed two training rounds to get used to activating target barrels, experiencing distractions, and seeing visual cues. Each stage was preceded by a brief training session to get subjects accustomed with the task.

4.3.1. Attention Restoration Stage

The attention restoration stage used a 10x2x2x3 study design with the following variables: **Cue**: which cue was seen, **Target Spacing**: narrow or wide barrel spacing (as described earlier), **Distraction Distance**: close or far placement of spherical distraction objects (distance from current target barrel), and **Distraction Breadth**: number of objects in the distraction (1, 3, or 5). Subjects performed 2 rounds of the restoration task, each with a different Target Spacing. In each round, the cue was used 6 times, each time with a different combination of Distraction Breadth and Distance.

Subjects were given the following instructions before each attention-restoration round: "Activate each barrel by looking at it for 1 second in order of ascending label. Look at all spheres when they appear, starting with the red sphere. Notify the proctor when you are ready to begin."

Subjects gazed at barrels in ascending label order, randomized according to the procedure noted in subsection 4.2, for 1 second each. While gaze dwelled on a barrel, its label turned light green. Once the target barrel had been gazed at for 1 second, it was "activated" and its label turned a darker shade of green. For every 2 or 3 activated barrels (random), a distraction appeared while the subject dwelled on a barrel. This happened after 0.5-1 seconds of dwell (random). Once the distraction task began, a visual cue was rendered to help guide them back. After activating a full set of 8 barrels, the barrel pattern was re-randomized and reset. This was done as many times as needed until 6 distraction events had occurred.

To handle the distraction objects, a subject had to look at all presented spheres, in an order, to remove the spheres. Distractions appeared accompanied by a "beep" sound, and the number of spheres was determined by Distraction Breadth. Spheres were placed on an arc such that they had equal distance from the

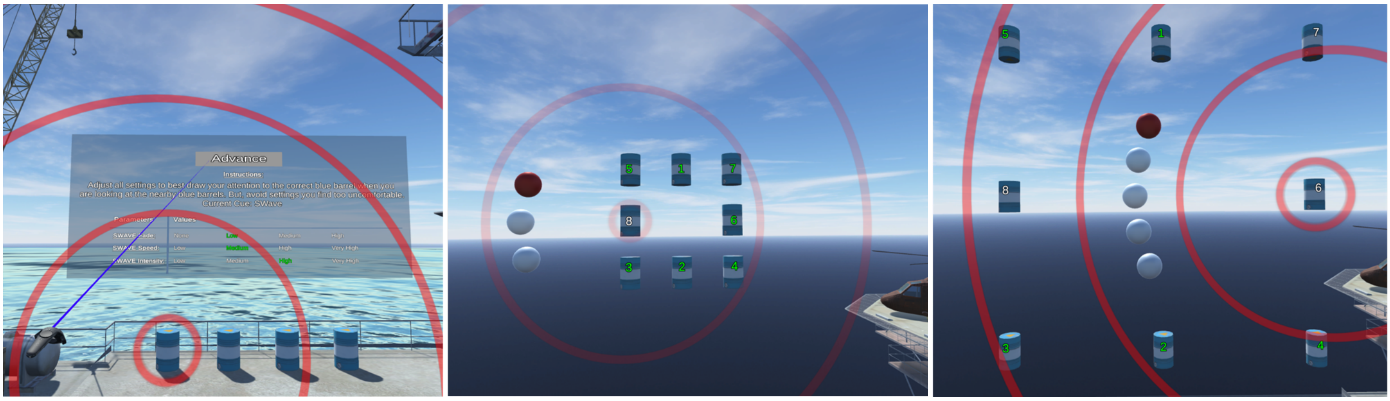

Fig. 2. Environment layouts used in experiments. Left: narrow target spacing during the tuning study. Center: wide target spacing with 5 far distractions during restoration stage. Right: narrow target spacing with 3 close distractions during restoration stage.

target barrel. A sphere at one end of the arc (chosen randomly) was colored red and others (if present) were initially white. The subject had to look at the red sphere, causing it to vanish, and the next closest sphere along the arc turned red, denoting that it was the new target. Each red sphere disappeared when it was gazed at for 3 graphics frames. The 3-frame threshold avoids activation by jitter or approximate moves over the distraction. The result was subjects performing a "sweeping" motion with their gaze across the arc of spheres.

Distraction Distance had 2 levels: close and far. Close and far spheres were placed at a distance of approximately 4.2 and 7.1 meters from the target barrel, respectively. Distraction Breadth had 3 levels: 1, 3, or 5 distraction spheres. For 1 sphere, the sphere was positioned either vertically/horizontally (for the middle barrels) or diagonally (for corner barrels) away from the barrels by the sphere distance. For 3 or 5 distractions, additional spheres were placed in an arc, rotated about the target with the first sphere. When Target Spacing was narrow, distractions appeared outside of the barrel layout; when it was wide, distractions appeared within the barrel layout (as seen in Figure 2).

Once the subject finished the distraction task, they returned their gaze to the target object, guided by the visual cue. Subjects completed trials with each combination of Distraction Distance and Distraction Breadth once per cue, thus each cue was used 6 times in a row. The order of combinations was randomized for each cue, the order of cues per attention-restoration round was randomized, and the order of Target Spacing was randomized.

4.3.2. Guidance Stage

The guidance stage used a 10x2x2 study design with the following independent variables: **Cue**: which cue was seen, **Target Spacing**: narrow or wide (as before), and **Mandated Dwell Time**: how long the subject must look at a target barrel before it is considered activated, either short (.125 second) or long (1 second). The short dwell was intended to be imperceptible, being just long enough to ensure that subjects glance at each barrel while moving at a fast pace with minimal thought. The 1-second pause was considered more consistent with prior studies in that it does not proceed immediately, but the task is also more generic than other studies in that it does not involve any longer or interactive tasks to be performed.

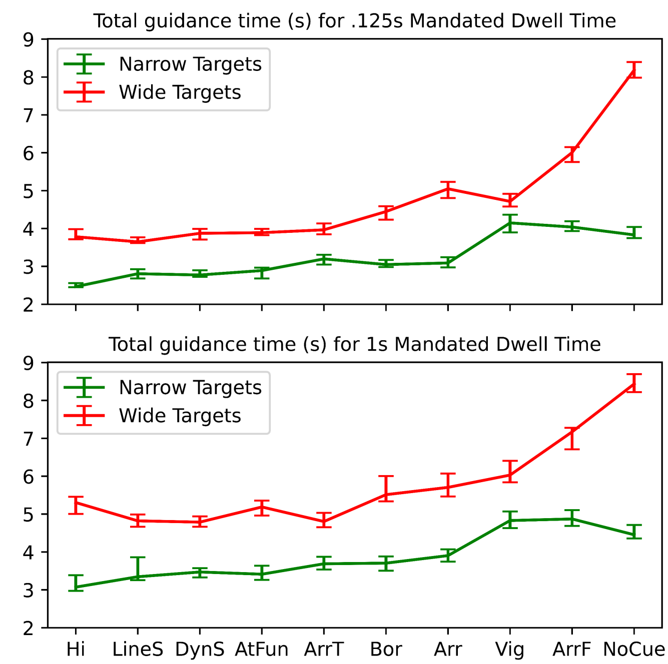

Fig. 3. Median guidance time for each cue at different levels of Target Spacing and Dwell Time. Error bars were computed by bootstrapping and show the range containing the middle 68 percent of 5000 median estimates, each being the median of 98 values (sampled with replacement). This is analogous to standard error when using parametric data.

Subjects were given the following instructions with in-VR text before each guidance round: "Activate each barrel by looking at it in order of ascending label. A cue will be present to guide you. Notify the proctor when you are ready to begin." As in the Attention Restoration stage, barrel labels were again randomized according to the procedure in subsection 4.2.

Subjects completed 4 rounds of the guidance task, each using a different combination of Target Spacing and Mandated Dwell Time. We consider the use of a cue to activate a full set of 8 barrels as a single trial, creating 10 trials per round (1 for each cue condition. No distractions occurred and the cue was active throughout each trial.

Table 2. Pairwise Wilcoxon comparisons for total guidance time between different cues. One pairwise test was performed for each combination of Target Spacing and Dwell Time levels, making a total of 4. Number and color show the number of significant differences found between the pairs ($p < .05$ after Holm correction). A negative (-) sign indicates that the row's cue had a lower (better) time than the column cue.

Cue	Hi	LineS	DynS	AtFun	ArrT	Bor	Arr	Vig	ArrF	NoCue
Hi	0	0	0	0	-2	-3	-3	-4	-4	-4
LineS	0	0	0	0	-1	-2	-3	-4	-4	-4
DynS	0	0	0	0	-1	-3	-3	-4	-4	-4
AtFun	0	0	0	0	-1	-1	-2	-4	-4	-4
ArrT	2	1	1	1	0	-2	-2	-4	-4	-4
Bor	3	2	3	1	2	0	-1	-3	-4	-4
Arr	3	3	3	2	2	1	0	-2	-4	-4
Vig	4	4	4	4	4	3	2	0	-1	-2
ArrF	4	4	4	4	4	4	4	1	0	-2
NoCue	4	4	4	4	4	4	4	2	2	0

4.4. *Questionnaires*

After each attention-restoration round or guidance trial, a usability questionnaire panel appeared in the scene asking the subject to rate the cue on a 7-point scale. It asked two questions: 1) "How well did the cue guide your attention to the correct barrel?" and 2) "How much do you like the cue?", with 1 being "not very well/much" and 7 being "very well/much." Subjects could select responses using ray-based interaction with the panel.

After each attention-restoration round, a ranking questionnaire appeared in the scene. 10 cue cards were presented, representing the 10 Cues. Subjects were asked to "Order the cue cards from worst to best in terms of how strongly they attracted your attention to the correct barrel in the previous trial." and "Order the cue cards from worst to best in terms of how much you like its appearance." Subjects ordered cards freely using ray-based interaction. When subjects pointed the ray at a card, the corresponding cue was displayed for 1 of 5 target barrels placed above the cue cards, for reference.

4.5. *Results*

4.5.1. *Metrics*

We consider cues that can quickly guide or restore attention while minimizing incorrect target seeking to be superior. To that end, we use time and target error as our primary metrics. For the attention restoration stage, the **lookback time** is the time between the user finishing the distraction task and shifting their gaze back to the correct target. The **lookback error count** is the number of incorrect targets the subject looked at before finding the correct one.

For the guidance stage, the **guidance condition time** is the total time taken to move through an 8-barrel sequence. Guidance time is additionally adjusted by subtracting the total mandated dwell time to enable more straightforward comparison between the two dwell time conditions. Total guidance time was further broken down into **movement time**, amount of time taken to move from one target to the next, and **dwell delay**, how long after the target switched that the subject lingered on the previous target. **Target error count** for guidance is the number of incorrect targets looked at across a trial. To avoid counting an erroneous intermediate target along a move to the correct

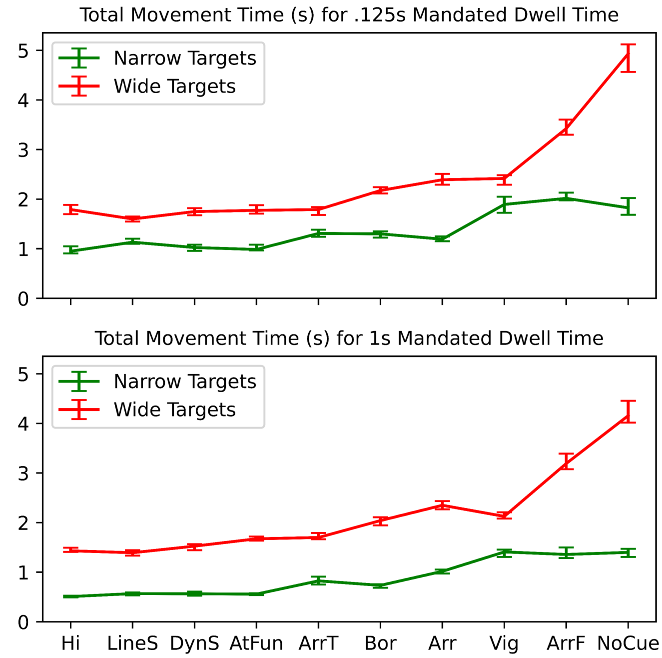

Fig. 4. Median time to move from one target to the next for each cue at different levels of Target Spacing and Dwell Time.

Table 3. Pairwise Wilcoxon comparisons between cues for movement time between targets. A negative (-) sign indicates less movement time for the row's cue.

Cue	Hi	LineS	DynS	AtFun	ArrT	Bor	Arr	Vig	ArrF	NoCue
Hi	0	0	0	0	-2	-4	-4	-4	-4	-4
LineS	0	0	0	-1	-2	-3	-3	-4	-4	-4
DynS	0	0	0	0	-3	-4	-3	-4	-4	-4
AtFun	0	1	0	0	-2	-4	-3	-4	-4	-4
ArrT	2	2	3	2	0	-2	-1	-4	-4	-4
Bor	4	3	4	4	2	0	0	-2	-4	-4
Arr	4	3	3	3	1	0	0	-2	-4	-4
Vig	4	4	4	4	4	2	2	0	-2	-2
ArrF	4	4	4	4	4	4	4	2	0	-2
NoCue	4	4	4	4	4	4	4	2	2	0

one, an incorrect target for both error metrics is only considered looked at if the user dwells on it for 100ms.

Subjective measures are also considered from the questionnaires given throughout the experiment. Secondary measures, such as total eye and head movements, were also recorded, and are reported in Appendix B.

Much of the data shows positive skew, with high-time and high-error outliers. So, analysis primarily uses Friedman tests to determine an overall effect caused by the independent variable, and pairwise Wilcoxon signed-rank tests in cases where there is an effect. Due to the large number of pairs for 10 visual cues, we correct p values using Holm corrections [36] to avoid familywise error inflation. While multiple independent variables are used in study, we focus primarily on differences between cues.

4.5.2. *Guidance Condition Time*

Total guidance time between cues with all levels of other variables included are summarized in Figure 3. A Friedman test showed that cue condition affected guidance time ($\chi^2(9) =$

472.15, $p < .001$). We performed pairwise tests separately within each of the 4 combinations of Target Spacing and Mandated Dwell Time. This allows us to compare the use of each cue in different scenarios. Table 2 summarizes results of pairwise Wilcoxon tests between the resulting 4 sets of data.

We suggest that results support putting the cues into 3 categories. Highlight, LineStrip, DynSWAVE, and AtFunnel all perform better than all other cues in at least 1 case, with their overall medians being comparable. ArrowTrail, Border, and Arrow perform worse than the previous group but better than the last 3. Vignette, ArrowField, and NoCue perform worse than the others. We postpone a fuller discussion comparing cues until Section 5 to first show all metrics.

Comparing levels of Target Spacing and Mandated Dwell Time, with all other values collapsed, gives us insight into the nature of the guidance task itself. Wide Target Spacing (median 5031ms) incurs significantly more time than Narrow (3472ms, $p < .001$). This is intuitively expected, as more space between targets will require more movement between targets and may decrease the chance that the next target will be seen in the subject's periphery. Long Mandated Dwell Time (median 4781ms) incurs significantly more time than Short (median 3791ms, $p < .001$). We suspect this may be due to longer time taken to react to new target switches after the eye has come to a rest on a target for a full second.

We see several interesting differences within individual cues when varying Target Spacing and Dwell Time. The general difference between wide and narrow Target Spacings is also shown when comparing within each cue, and in each case the guidance time is again higher with wide target spacing ($p < .001$). Additionally, some cues appear to have larger gaps between their performance with wide or narrow targets, such as Highlight, Arrow, and NoCue.

4.5.3. Guidance Time Breakdown

We consider and measure time for two primary actions the subject performs during the guidance task: movement and dwell. Movement is the time between the subject leaving one target and reaching the next correct target. This includes saccadic movements and the search process to reach the correct target, including movements towards and dwell on incorrect targets. Dwell includes mandated dwell time on the correct target and the time from when the target switches until the subject moves their eyes off the old target (here named dwell delay). We consider dwell delay an indication of how quickly the cue triggers a subject's response, with lower delay seen as positive.

We considered further dividing the total dwell time to include an error metric when subjects look off of the correct target, look at an incorrect target, then back at the correct one. While the times were non-zero, and showed differences between Target Spacing and Dwell Time levels ($p < .001$ for both), differences were not shown between cues themselves. As such, we exclude this analysis.

As seen in Figure 4, movement time shows a similar pattern to total guidance time in shape between cues and pairwise comparison results. Once again, Wide Target Spacing (median 2051.5ms) incurs more movement time than Narrow (1050ms, $p < .001$). Interestingly, results for Mandated Dwell Time are reversed, with Long Dwell Time (1391ms) incurring less than Short (1724ms, $p < .001$). Possibly, the longer dwell time allowed subjects to see the next target in their periphery and prepare to move towards it.

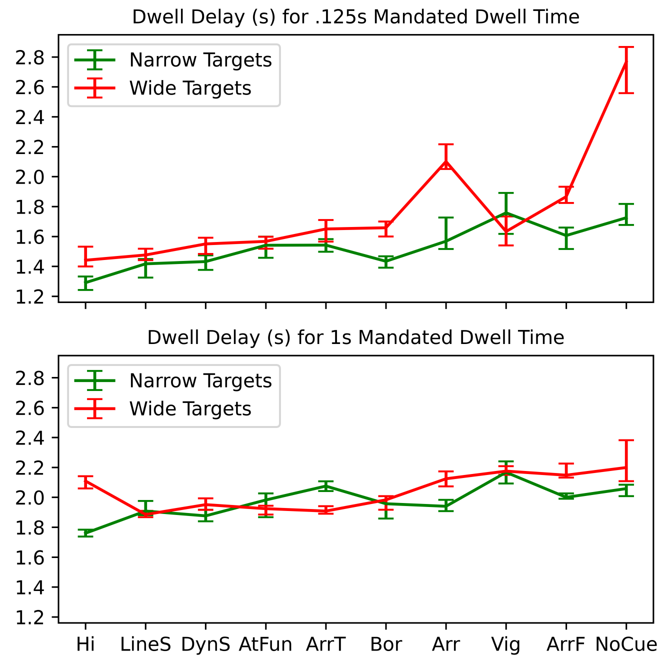

Fig. 5. Median delay to move gaze off an old target for each cue at different levels of Target Spacing and Dwell Time.

Table 4. Pairwise Wilcoxon comparisons between cues for dwell delay. A negative (-) sign indicates less delay for the row's cue.

Cue	Hi	LineS	DynS	AtFun	ArrT	Bor	Arr	Vig	ArrF	NoCue
Hi	0	-2	0	-2	-3	-2	-3	-3	-3	-4
LineS	2	0	0	0	-3	-2	-2	-4	-3	-4
DynS	0	0	0	0	-2	0	-2	-3	-4	-4
AtFun	2	0	0	0	0	0	-2	-3	-2	-2
ArrT	3	3	2	0	0	0	-3	1	-2	-2
Bor	2	2	0	0	0	0	-2	-1	-2	-2
Arr	3	2	2	2	3	2	0	-3	-1	-3
Vig	3	4	3	3	-1	1	3	0	1	-1
ArrF	3	3	4	2	2	2	1	-1	0	-1
NoCue	4	4	4	2	2	2	3	1	1	0

Dwell delay (seen in Figure 5) shows slightly different patterns. With the exception of Highlight, cues that immediately appear near the subject's focal point appear to perform best. Highlight shows poor performance with Wide targets and Long dwell time, with a similar median to NoCue, implying that since the Highlight can appear outside their field of view they may have not noticed the target switch. Arrow shows poor performance with Wide targets and Short dwell time, possibly because the arrow is drawn in the periphery between the gaze point and target, and thus subjects take longer to respond. NoCue shows poor performance in the same case.

4.5.4. Guidance Error Counts

Errors were counted if a subject's gaze dwelled on an incorrect target for longer than 100ms, and were tallied across the 8-target trial. Results per cue are shown in Figure 6. The pat-

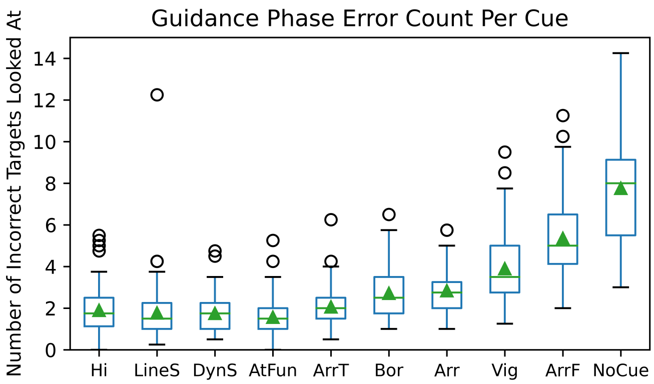

Fig. 6. Number of erroneous targets gazed at throughout an 8-target guidance trial for each cue.

Table 5. Pairwise Wilcoxon comparisons between cues for restoration lookback time. A negative (-) sign indicates less lookback time for the row's cue.

Cue	Hi	LineS	DynS	AtFun	ArrT	Bor	Arr	Vig	ArrF	NoCue
Hi	0	-3	-1	-7	-5	-4	-6	-4	-3	0
LineS	3	0	0	-4	0	0	0	-1	0	0
DynS	1	0	0	-7	0	-1	-3	-2	-1	0
AtFun	7	4	7	0	0	2	1	0	2	5
ArrT	5	0	0	0	0	0	0	0	0	0
Bor	4	0	1	-2	0	0	0	0	0	1
Arr	6	0	3	-1	0	0	0	-1	0	0
Vig	4	1	2	0	0	0	1	0	0	0
ArrF	3	0	1	-2	0	0	0	0	0	0
NoCue	0	0	0	-5	0	-1	0	0	0	0

tern of results generally aligns with overall guidance time, i.e., if a cue incurs more time, it is more likely that there were errors. Interestingly, Wide Target Spacing (median 1 error) incurs significantly fewer errors than Narrow (median 3 errors, $p < .001$), a reverse of the trend seen in overall guidance time. We suspect it was less likely to reach a wrong target when targets were farther apart and fewer targets are in view at one time.

4.5.5. Attention Restoration Lookback Time

For the restoration task, median lookback times for all cues at every combination of the other 3 variables are summarized in Figure 7, with Wilcoxon pairwise comparisons summarized in Table 5. The most notable result is that median times for the cues do not follow the same patterns as they did in the guidance stage. The primary differences appear to lie with Line Strip, Attention Funnel, Arrow Trail, and No Cue.

While No Cue performed significantly worse than most other cues during guidance, here it performed better than Vignette ($p = .024$) and Attention Funnel ($p = .015$), was not shown to be significantly worse than any other cue, and had the third lowest median lookback time. We suspect this relates to the nature of the task; subjects had previous knowledge of the target (in contrast to a guidance task) and knew they had to return to the target once they cleared the distraction spheres. This is analogous to a quick distraction such as glancing to a tablet to dismiss a notification. It may be that some added visuals like the Attention Funnel have a negative effect of distracting the user from directly returning to a known target.

While Line Strip, Attention Funnel, and Arrow Trail were among the better cues for guidance, here they fared significantly worse. Line Strip incurred higher lookback times than Highlight ($p = .02$). Arrow Trail did worse than Highlight ($p < .001$) and DynSWAVE ($p = .016$). Attention Funnel did worse than Highlight ($p < .001$), DynSWAVE ($p < .001$) and No Cue ($p = .015$).

Regarding the 3 other variables, of primary interest is the Distraction Breadth. We initially assumed a larger number of distraction spheres would incur higher lookback times. However, the median time for 1 distraction (325ms) is higher than 3 distractions (267ms, $p < .001$) and 5 (267ms, $p < .001$). The medians for 3 and 5 distractions do not appear to differ ($p = .796$).

Regarding Distraction Distance, we expected that far distractions (median 308ms) would incur more time than close distractions (267ms, $p < .001$). Visual inspection of Figure 7 appears to also suggest there may be an interaction between cue and Distraction Distance, with cues such as Arrow showing possible wider variations between close and far than cues like DynSWAVE. In this sense, choice of cue may matter more for farther distractions.

4.5.6. Attention Restoration Lookback Error Count

Similar to the guidance stage, lookback errors were counted if a subject's gaze dwelled on an incorrect target for longer than 100 ms before finding the correct target. Counts of errors per trial are shown in Figure 8. Resulting distributions resemble a Poisson distribution (with low λ), with heavy skew towards 0 errors, thus we compared cues using a Poisson test. Pairwise comparison results are shown in Table 6.

Unlike in the guidance stage, lookback errors do not clearly follow the same pattern as lookback time. While Highlight is one of the best performing for lookback time, it has the highest maximum error count and only shows significantly fewer errors than ArrowField ($p < .001$). While Attention Funnel performed worse than most cues in lookback time, it did not show significantly more errors than any cue, and showed significantly fewer than Vignette ($p = .005$), Arrow Field ($p < .001$), and No Cue ($p = .017$).

Perhaps most importantly, while No Cue outperformed Attention Funnel and Border in lookback time, it resulted in more errors than Line Strip ($p = .002$), DynSWAVE ($p < .001$), Attention Funnel ($p = .017$), and Arrow Trail ($p = .008$). Though subjects were able to quickly find the correct target without a visual cue, they were more likely to look back at an incorrect target first, suggesting the cues were still useful for targeting.

4.5.7. Subjective Ranking

Subject rankings given at the end of each attention restoration round were numerically coded (1 is low, 10 is high) and averaged between a subject's ranking for wide and narrow target spacing rounds. Results are summarized in Figure 9. Likert-like ratings were also collected at the end of each use of a cue, but results do not differ substantially from rankings and are thus not presented.

Pairwise comparisons, summarized in Table 7, follow a similar pattern from guidance results. Given the number of cues, a clear final ranking is not possible, but considering the number

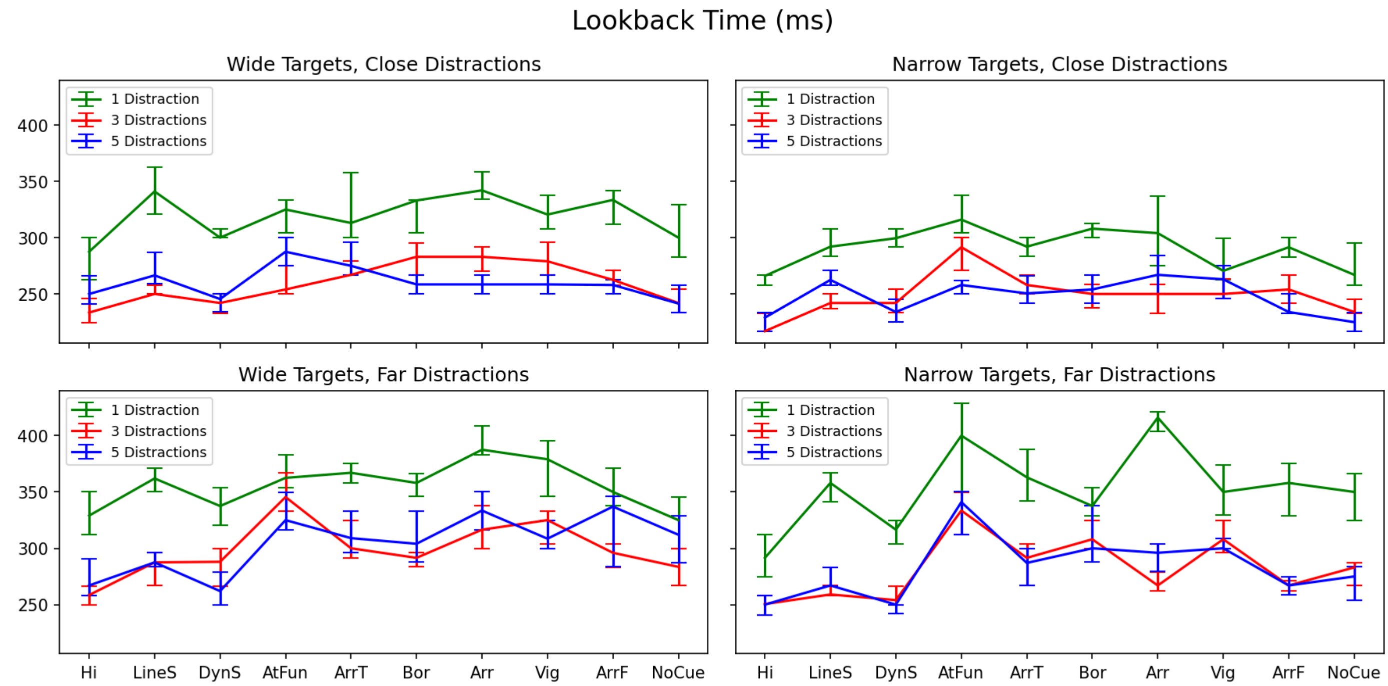

Fig. 7. Median time to look back from distractions to correct target for each cue at different levels of Distraction Breadth, Distraction Distance, and Target Spacing. Error bars were computed by bootstrapping and show the range containing the middle 68 percent of 5000 median estimates, each being the median of 67 values (sampled with replacement).

Table 6. Pairwise comparisons between cues for lookback errors. A negative (-) sign indicates that row's cue was ranked lower than the column cue.

Cue	Hi	LineS	DynS	AtFun	ArrT	Bor	Arr	Vig	ArrF	NoCue
Hi	0	0	0	0	0	0	0	0	-1	0
LineS	0	0	0	0	0	0	-1	-1	-1	-1
DynS	0	0	0	0	0	0	-1	-1	-1	-1
AtFun	0	0	0	0	0	0	0	-1	-1	-1
ArrT	0	0	0	0	0	0	-1	-1	-1	-1
Bor	0	0	0	0	0	0	0	-1	-1	0
Arr	0	1	1	0	1	0	0	0	-1	0
Vig	0	1	1	1	1	1	0	0	-1	0
ArrF	1	1	1	1	1	1	1	0	0	0
NoCue	0	1	1	1	1	0	0	0	0	0

of times a cue is significantly better or worse than another, we see a rough ranking into 6 groups. ArrowTrail, AtFunnel, and LineStrip are ranked at the top, followed by DynSWAVE, then Arrow and Highlight, then Border, then Vignette, then ArrowField and NoCue.

Interestingly, despite Highlight's good performance in both stages, it is ranked only among the middle of the cues. Conversely, NoCue performed well in attention restoration, but was ranked low. Looking at the top ranked cues, they all render a visual component anchored near the user's gaze point and leading directly to the target.

5. Discussion

In the following discussion, we primarily focus on differences and similarity between the visual cues based on performance. Cues that are visually similar are grouped to better compare/contrast them.

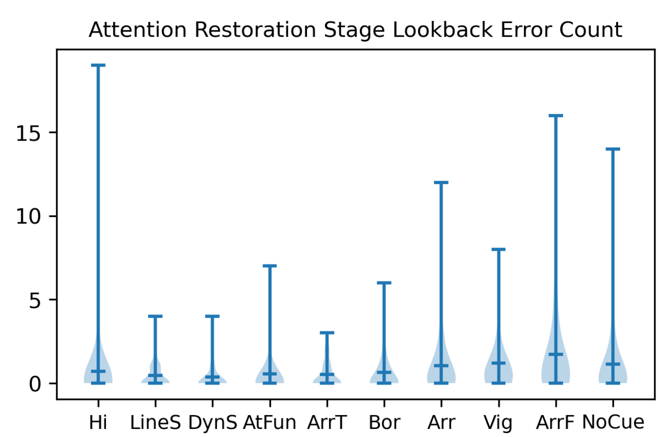

Fig. 8. Violin plot of the number of erroneous targets looked at when looking back after a distraction event. Middle lines denote means; top and bottom lines denote maximum and minimums.

Table 7. Pairwise comparisons between cues for subjective ranking. A negative (-) sign indicates that row's cue was ranked lower than the column cue.

Cue	Hi	LineS	DynS	AtFun	ArrT	Bor	Arr	Vig	ArrF	NoCue
Hi	0	-1	0	-1	-1	0	0	1	1	1
LineS	1	0	0	0	0	1	1	1	1	1
DynS	0	0	0	0	0	1	0	1	1	1
AtFun	1	0	0	0	0	1	1	1	1	1
ArrT	1	0	0	0	0	1	1	1	1	1
Bor	0	-1	-1	-1	-1	0	0	1	1	1
Arr	0	-1	0	-1	-1	0	0	1	1	1
Vig	-1	-1	-1	-1	-1	-1	-1	0	1	1
ArrF	-1	-1	-1	-1	-1	-1	-1	-1	0	0
NoCue	-1	-1	-1	-1	-1	-1	-1	-1	0	0

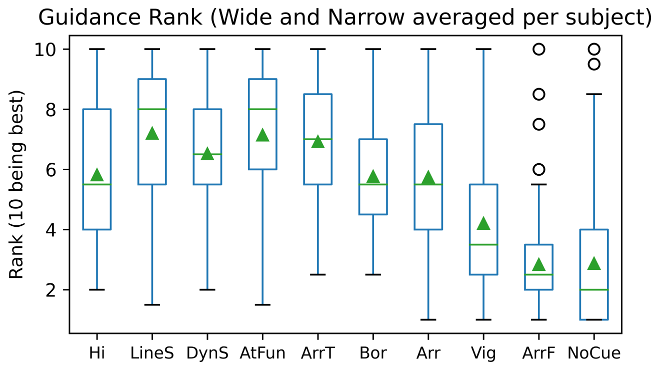

Fig. 9. Subjective rankings for each cue, with 10 being best. Rankings were provided for cues during wide and narrow Target Spacing, and were averaged per subject.

5.1. Visual Cues Discussion

5.1.1. Highlight

Highlight is one of the best-performing cues. For attention restoration (Table 5), it had better times than all but 2 other cues. For attention guidance (Table 2) it is again one of the strongest, showing better performance than all but 3 cues in at least 1 case. Given the short Dwell Delay with narrow targets (on average less than 200 ms), we suspect the Highlight's quick change in color may trigger a preattentive reflexive saccade [37]; subjects noted the effect occasionally making their eyes move to it automatically before they were consciously aware of it.

Important to Highlight's design is that the effect is placed on the target itself. A benefit to this is potentially that it draws attention to the target directly rather than indirectly through a visual effect component at another location. A possible drawback is if the user does not have the target in view, there should be no effect. This may be why it performs very well in lookback time (Table 5) but not as well in lookback errors (Table 6). We also expected Highlight to perform poorly with wide Target Spacing, but the difference between narrow and wide does not appear much larger than the difference for other cues.

We do see in the guidance task (Figure 3) a large gap between narrow and wide performance with long Mandated Dwell Time, with wide on median taking about 2 seconds longer to complete. We expected subjects did some broader search before catching the Highlight effect in their periphery, then moved quickly towards the target. Were this the case, we would expect to see higher Movement Time in the wide case, but this is not shown in data (Figure 4). Instead, we see the added time in the long-dwell/wide-targets case coming from dwell delay. Thus we suspect that the lack of imagery popping immediately into view (unlike cues like Line Strip) led to subjects relying on less accurate internal time counting or the more subtle cue of the barrel's number switching from light green to dark green to know when to gaze away from the target, but once they did, they still found the target very quickly.

5.1.2. DynSWAVE and Border

These cues are grouped because they both visualize a ring(s) around the target. DynSWAVE is another top-performing cue. In the restoration task, its lookback time was lower than all but NoCue, Highlight, and LineStrip. In the guidance task it is also one of the strongest, showing a low error count and better performance than many in guidance time (Table 2). The design of the cue makes it immediately noticeable, which likely contributed to resulting in lower dwell delay than other cues with more subtle effects (Table 4). We expected a longer search time, given that cues with animated speeds have sometimes been shown to induce eye movement at that speed (e.g. FlyingARrow in [24]). However, DynSWAVE outperformed six other conditions in movement time (Table 3), implying that effect was not present in this task.

DynSWAVE outperformed Border in several cases while Border never performed better than DynSWAVE (Table 2). This is most noticeable in the guidance task's Movement Time where it outperforms Border in all 4 conditions. According to subject feedback, while the arc of Border's ring helped guide users in the right direction, the indirect nature of its pointing led to some "hot and cold" aspects to the searching.

Despite performing well, there are concerns about DynSWAVE's design. It is visually large and obstructive. This likely helps the user notice it immediately, leading to lower dwell delay. But it may also clutter parts of the environment the user wishes to inspect (e.g. scanning the environment during a training explanation). The obstructive nature also makes it impractical for extending to multiple targets, and may be inappropriate for some tasks. Additionally, all of our targets were far away and generally in front of the user; the cue may have a stranger appearance if the distance to the target is less than the user's height as the rings could appear to pass through their body.

5.1.3. Line Strip and Attention Funnel

These cues are grouped together because they are aesthetically different versions of the same concept: visualizing a curve path from a short distance in front of the eyes towards the target. In the restoration task (Figure 7), Line Strip sits somewhere in the middle while Attention Funnel has the worst median lookback time in several cases. We also see that the lookback time for Attention Funnel is worse than Line Strip in 4/12 cases as well (Table 5). Despite this, both cues resulted in minimal lookback errors (Table 6). Had the cues resulted in high errors, we might infer that they were misleading. Instead, we infer that they were distracting, and led the subject to look directly at the cue first, then the target.

Both cues performed much better in the guidance task than in restoration, scoring much closer to each other and showing better times than 6 of the other cues in at least 1 condition (Table 2). Both also resulted in low guidance error (Figure 6). Additionally, they (along with the Arrow Trail) have the highest subjective ranking.

We assumed the cues would be relatively self-explanatory, and so there was no "tutorial stage" for cue usage, but several subjects seemed confused when the Funnel appeared. We consider that because there is no direct connection between the rings of Attention Funnel, subjects did not have an intuitive sense of how to follow it. Line Strip, however, shows the curve minimally, directly, and continuously, which may explain why it performed better than the Funnel.

We suggest the difference in performance between guidance and restoration is due to differences in the tasks themselves. While the cues rendering directly in front of the user appears to be useful in guidance tasks, it may actually be a hindrance and distraction in restoration tasks.

Because Line Strip and Attention Funnel render a visual component directly at the gaze point, we consider that these may benefit from the inclusion of eye tracking more than most of our studied cues. Without eye-tracking, cues would have to be rendered at at an approximated head gaze position, creating a potential offset between gaze and cue which may trigger extra motion or reflexive motion as the user is drawn first to the cue, then to the target. This is potentially evidenced by long guidance times with the Arrow Field cue, in which we expect users had to look for the arrows, then process the direction in which they were pointing.

5.1.4. Arrow and Arrow Trail

These cues are grouped together because they both place an arrow on an arc with head-to-arc distance equal to head-to-target distance.

In the restoration task, Arrow Trail only performs better than Attention Funnel and worse than Highlight and DynSWAVE (Table 5), thus ranking around the middle. This is reinforced in the guidance task, where, for total guidance time, Arrow Trail performs better than Arrow, Border, Vignette, and Arrow Field, but worse than Highlight, DynSWAVE, Line Strip, and Attention Funnel (Table 2).

In the restoration task, Arrow is only shown to perform worse than DynSWAVE and Highlight. In guidance, it outperforms No Cue, Vignette, and Arrow Field, but performs worse than all others at least once.

The Arrow's problem may relate to the its placement between the gaze and the target. This offset was chosen in the tuning study and is a compromise between a near-target arrow that may be out of view and a gaze-placed arrow that must be first interpreted in-place to understand its direction. The user may "chase" an arrow in the periphery. In the restoration task (Figure 7), there is a spike in lookback time for the 1-far-distraction narrow-targets case; since the distraction was farther away, the subject likely took more time to notice it was rendered. This spike happens to a lesser degree as well in the wide targets case, and was likely smaller due to the arrow's offset (15% for wide compared to 45% for narrow) leaving it closer to the eye and more immediately noticeable.

During guidance (Figure 3) we also see poor performance times in the wide targets cases, which appears to be worse with a short mandated dwell time. This is also replicated in movement time, and to a larger extent in dwell delay with a Short Mandated Dwell Time. This suggests that Arrow takes longer both to be noticed and to guide users. Longer dwell delay is likely caused by its peripheral design; longer movement time may be related to its tendency to incur error.

Because the arrow does not directly connect to the target, if the eye is far away from the correct target and there is an incorrect target along the path, it could appear like the arrow is pointing to the incorrect target. This is supported by Arrow showing higher error counts than the ArrowTrail and LineStrip (which do directly connect to the target) in both restoration (Figure 6) and guidance (Table 6).

5.1.5. Vignette

Vignette had one of the highest median lookback times in the restoration task (Figure 7); it was shown to be worse than Highlight, DynSWAVE, and No Cue. This holds true for lookback error as well, where it performed worse than most cues (Table 6). The trend continued in the guidance task, where for total guidance time it was only shown to perform better than No Cue and Arrow Field. In closing subject interviews, they often said they did not like how obtrusive it was and that it was too indirect to see exactly where it was pointing.

Given its large and obtrusive nature, we expected a positive aspect of the cue would be subjects immediately noticing it, thus helping to lower guidance and lookback times. However, this appeared to not be the case, as Vignette incurred more dwell delay time than all conditions aside from No Cue and Arrow Trail (Table 4). Additionally, like Border, Vignette uses a "circular" style of cue that indirectly suggests a broad potential of target locations instead of a precise targeted area, potentially evidenced by Vignette incurring more movement time (Table 3) and error (Figure 6) than most cues during the guidance task.

5.1.6. Arrow Field

Arrow Field has the second highest median lookback time of all cues (Figure 7), and the worst overall performance of all cues in the guidance task (Table 2). Subject interviews revealed many thought it hurt more than it helped and that they didn't like having to look elsewhere at multiple arrows to find where they were pointing. Two subjects asked if it was a misleading cue designed to make the task harder.

We suspect poor performance was due to two primary factors. First, the arrows are not placed on a path between the gaze point and target, likely meaning that to find the target, subjects had to first move their gaze off the path to the target, then interpret the direction, then move to the target. This is supported by high lookback (Figure 7) and guidance (Figure 3) times, and supports the idea that a cue's position may better communicate direction than its shape. Second, we suspect something about the spherical coordinates on which it was placed made it not always obvious where any 1 arrow was pointing; if subjects didn't know to follow the arrow's path on a sphere it might look like it was pointing to a completely different target. This is supported by Arrow Field incurring more errors during both restoration (Table 6) and guidance tasks (Figure 6).

5.1.7. No Cue

No Cue performed surprisingly well during the restoration task: it incurred the third lowest median lookback time, significantly lower than Vignette and Attention Funnel (Table 5). However, while subjects found the correct target quickly, lookback error revealed that they were more likely to find incorrect targets along the way (Table 6). We suggest this is because subjects could remember roughly where the target they needed to look back to was, but did not have a visual pointing out exactly which target was correct.

Results for No Cue only worsened for the guidance task where it incurred more total time, movement time, dwell delay, and errors than every other cue in at least one case (Table 2), with results being particularly bad in the case of wide targets. This is likely because wide target spacing makes it more difficult to both notice that the target has switched and quickly see which barrel is the correct target within periphery. We suggest the difference between performance in these tasks is due to memory of target position for restoration.

5.2. Restoration vs. Guidance

Judging from cue performance we see 2 main differences between the 2 tasks: moving from restoration to guidance, Attention Funnel and Line Strip perform much better and No Cue performs much worse. No Cue worsening likely points to a greater difference between the 2 tasks.

In restoration, users already know where the target they need to return to is, and that they need to return to it. Critical here is that the distraction is brief and forced, analogous to quickly addressing a notification in a training simulation. The user is not necessarily mentally distracted from the task as they might be during a longer or more complex task. Since the user can keep in mind that they need to return, and they already know where the target is, a cue might actually distract them from getting back to it, as evidenced by the performance of cues like Attention Funnel and Arrow Field. When evaluating the effect of Distraction Distance on No Cue, we also see particularly large time gaps between near and far distractions (Figure 7), suggesting that when the subjects gazed far enough away from the target that it gets out of view, it may take longer to find again without assistance. Higher error counts also suggest that subjects were more likely to look at incorrect targets on the path to the correct one, a problem which may worsen with a higher number of denser targets.

In guidance, the user does not know where the target will be and has to find it, explaining No Cue's poor performance. Looking at total guidance time, when targets are narrow the user can at least get a general sense for where the next target might be (presumably using peripheral vision), so the performance isn't dramatically worse than the worst cues, but when the targets are wide this worsens by several seconds.

Ultimately, if the situation calls for a user with no knowledge of where they're supposed to look to find a new target, a cue should definitely be used. If the situation calls for a user with previous knowledge of the target position to be guided back to the target, their own memory might be useful enough, but the lack of visuals may not snap them out of whatever distraction they're currently in, or may lead them to incorrect targets first. The difference in cue effectiveness patterns between the two task types may point to the existence of other patterns for other, unconsidered tasks. More specific distraction situations may need to be studied for more targeted guidance.

5.3. Distraction Breadth and Distance

Distraction breadth referred to 1, 3, or 5 spherical objects in the distraction. Our initial expectation was that increased breadth would increase lookback time, as the added movement and time could cause some forgetting about target position. Figure 7 appears to disprove this: most cues have a higher median lookback time for 1 distraction than for 3 and 5. Also, breadths 3 and 5 tend to have similar lookback times. This is consistent with pairwise tests, for which there is a difference between 1-3 and 1-5 pairs, but not between 3-5.

This may have more to do with the inertia or control of eye movement than the cue. We speculate the following: with 1 distraction, the subject moves to the distraction, takes a short pause to confirm that they've looked at the distraction and check if other distractions are present, then reverses direction 180 degrees to move back to the target. With 3 or 5 distractions the subject can move their eye over to the distraction space, pause to recognize that they have reached it and that there are additional distractions, then when they complete the sweep they are immediately aware that they can return and only have to change direction by 90 degrees.

Distraction distance varied the distance from the target (close or far). This varied the needed eye movement distance and put the target further out of field of attention.

Wilcoxon tests showed that close and far distractions differed with all other variables collapsed, and visual inspection suggests more distant distractions produced higher median lookback time in every case (Figure 7). While this is not surprising in general, we do visually infer some interaction between Distraction Distance and other variables besides cue. For example, Arrow had a particularly large gap with 1 distraction and narrow targets and Attention Funnel had a large gap with 3 distractions/wide targets and 5 distractions/narrow targets, while cues like Highlight and DynSWAVE have smaller gaps throughout. This may suggest that some cues are more subject to ambiguity in certain cases when the target goes out of field of attention.

5.4. Target Spacing Discussion

Target spacing was the spread of target barrels: narrow or wide. When narrow, all barrels could generally be in view at once, whereas with wide targets generally only 2-3 other targets were in the field of attention at once. While the wider targets also increase the amount of needed eye movement, we believe it is this periphery effect that has the largest effect on performance. It was our expectation that conditions using a narrow target layout would result in faster times than those with a wide layout.

In Figure 7 it appears this is not always true, as the medians sometimes overlap, though it is still often the case for cues that don't directly point to the target. Despite this, the wilcoxon test between narrow and wide did show differences ($p < .001$) though the effect is likely not large. However, looking at Figure 3, we see the effect appears much larger in the guidance task; there was a general gap between most cues, and a larger gap in cues that do not directly connect to the target. Looking at the breakdown of guidance time, it appears that most of this difference was in movement time, and not dwell delay. The larger gap in time for the guidance task compared to restoration may be due to the task trial generally taking longer (thus allowing time for gaps to become more pronounced) or because of some inherent difference in the task.

6. Conclusion

When discussing a preferred cue, we provide 3 main recommendations. If it is possible to include in the environment design, a well designed Highlight cue can provide quick guidance or restoration as long as the targets are within view, and is usable with multiple targets. If there is a singular out-of-view target, DynSWAVE will quickly grab the user's attention in either a restoration or guidance task. If DynSWAVE is too obstructive for the application environment or task, LineStrip along a curve will provide quick guidance and may be usable with multiple targets, but may not be appropriate for attention restoration. In fact, a designer may opt to show no cue for a restoration task unless they can confirm the user is actually distracted in a way that affects memory of prior visual targets.

Our work also gives insight into the nature of tasks used for these studies. Many prior works have focused on guidance-based tasks when evaluating cues. Results here show that those results are not directly applicable to an attention restoration task. Given the prevalence of distraction in VR environments, more research may need to be done specifically into restoration tasks.

Limitations and Future Work. Although we accounted for many variables within our restoration and guidance tasks, cues can still be tested in a variety of other conditions to see if they work in practical applications. For example, all of our targets were still and generally in front of the user, and the environment was static and designed with a small palette of colors. Distractions were not spontaneous, as the study needed to present many distraction events in a limited time, so the extent to which cues grab attention for an occasional or spontaneous distraction is not assessed.

Future work will attempt to overcome these limitations. By using a more natural educational or training environment with a real-world task, we can study the effects of cues in a more realistic setting. Targets can have more variable locations to test cues' abilities to guide attention to targets behind the user, and the environment can be more visually varied to test their ability to stand out to the user. A more sophisticated distraction simulator can allow for varied and controlled distractions that further guarantee a user will actually be distracted (e.g. unprompted noises, simulated phone notifications, etc...). Cues can be combined to take advantage of their strengths in given circumstances. This will give more insight into attention restoration as a task and how cues respond to more widely varying conditions. Further insight into studies of these tasks could also breakdown how subjects spend their time during trials and general eye movement patterns to see if those differ between cues as well.

Acknowledgments

This material is based upon work supported by the National Science Foundation under Grant No. 1815976.

References

[1] Wang, P, Wu, P, Wang, J, Chi, HL, Wang, X. A critical review of the use of virtual reality in construction engineering education and training. International Journal of Environmental Research and Public Health 2018;15(6). URL: https://www.mdpi.com/1660-4601/15/6/1204. doi:10.3390/ijerph15061204.

[2] Li, Y, Seo, BK, Kim, K. Exploring industrial uses of virtually altering the physical world. In: 2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). 2023, p. 434–437. doi:10.1109/VRW58643.2023.00094.

[3] Kang, Z, Jeon, J, Salehi, S. Eye tracking data analytics in virtual reality training: Application in deepwater horizon oil drilling operation. Proceedings of the Human Factors and Ergonomics Society Annual Meeting 2020;64(1):821–825. doi:10.1177/1071181320641191.

[4] dos Santos, IHF, Soares, LP, Carvalho, F, Raposo, A. A collaborative virtual reality oil and gas workflow. International Journal of Virtual Reality 2012;11(1):1–13. URL: https://ijvr.eu/article/view/2832. doi:10.20870/IJVR.2012.11.1.2832.

[5] Bozzi, LOS, Samson, KDG, Tadeja, S, Pattinson, S, Bohné, T. Towards augmented reality guiding systems: An engineering design of an immersive system for complex 3d printing repair process. In: 2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). 2023, p. 384–389. doi:10.1109/VRW58643.2023.00084.

[6] Albawaneh, A, Agnihothram, V, Wu, J, Singla, G, Kim, H. Augmented reality for warehouse: Aid system for foreign workers. In: 2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). 2023, p. 432–433. doi:10.1109/VRW58643.2023.00093.

[7] Cai, J, Yang, L, Zhang, Y, Cai, H. Estimating the visual attention of construction workers from head pose using convolutional neural network-based multi-task learning. In: Construction Research Congress 2020. ????, p. 116–124. doi:10.1061/9780784482865.013.

[8] Kim, S, Nussbaum, MA, Gabbard, JL. Influences of augmented reality head-worn display type and user interface design on performance and usability in simulated warehouse order picking. Applied Ergonomics 2019;74:186–193. URL: https://www.sciencedirect.com/science/article/pii/S0003687018303387. doi:https://doi.org/10.1016/j.apergo.2018.08.026.

[9] Binetti, N, Wu, L, Chen, S, Kruijff, E, Julier, S, Brumby, DP. Using visual and auditory cues to locate out-of-view objects in head-mounted augmented reality. Displays 2021;69. URL: http://dx.doi.org/10.1016/j.displa.xxxx.xx.xxx. doi:doi.org/10.1016/j.displa.2021.102032.

[10] Bork, F, Schnelzer, C, Eck, U, Navab, N. Towards efficient visual guidance in limited field-of-view head-mounted displays. IEEE Transactions on Visualization and Computer Graphics 2018;24:2983–2992. doi:10.1109/TVCG.2018.2868584.

[11] Gruenefeld, U, Lange, D, Hammer, L, Boll, S, Heuten, W. Flyingarrow: Pointing towards out-of-view objects on augmented reality devices. Association for Computing Machinery, Inc. ISBN 9781450357654; 2018,doi:10.1145/3205873.3205881.

[12] Lange, D, Stratmann, TC, Gruenefeld, U, Boll, S. Hivefive: Immersion preserving attention guidance in virtual reality. Association for Computing Machinery. ISBN 9781450367080; 2020,doi:10.1145/3313831.3376803.

[13] Nielsen, LT, Møller, MB, Hartmeyer, SD, Ljung, TC, Nilsson, NC, Nordahl, R, et al. Missing the point: An exploration of how to guide users' attention during cinematic virtual reality. vol. 02-04-November-2016. Association for Computing Machinery. ISBN 9781450344913; 2016, p. 229–232. doi:10.1145/2993369.2993405.

[14] Renner, P, Pfeiffer, T. Attention guiding techniques using peripheral vision and eye tracking for feedback in augmented-reality-based assistance systems. Institute of Electrical and Electronics Engineers Inc. ISBN 9781509067169; 2017, p. 186–194. doi:10.1109/3DUI.2017.7893338.

[15] Woodworth, JW, Yoshimura, A, Lipari, NG, Borst, CW. Design and evaluation of visual cues for restoring and guiding visual attention in eye-tracked vr. In: 2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW). 2023, p. 442–450. doi:10.1109/VRW58643.2023.00096.

[16] Bailey, R, McNamara, A, Sudarsanam, N, Grimm, C. Subtle gaze direction. ACM Transactions on Graphics 2009;28:1–14. doi:10.1145/1559755.1559757.
[17] Biocca, F, Tang, A, Owen, C, Xiao, F. Attention funnel: Omnidirectional 3d cursor for mobile augmented reality platforms. 2006, p. 1115–1122. URL: http://mindlab.orghttp://metlab.cse.msu.edu.
[18] Grogorick, S, Albuquerque, G, Magnor, M. Comparing unobtrusive gaze guiding stimuli in head-mounted displays. 2018, p. 2805–2809. URL: https://www.smivision.com/wp-content/uploads/.
[19] Gruenefeld, U, Ennenga, D, Ali, AE, Heuten, W, Boll, S. Eyesee360: Designing a visualization technique for out-of-view objects in head-mounted augmented reality. Association for Computing Machinery, Inc. ISBN 9781450354868; 2017, p. 109–118. doi:10.1145/3131277.3132175.
[20] Gruenefeld, U, Ali, AE, Boll, S, Heuten, W. Beyond halo and wedge: Visualizing out-of-view objects on head-mounted virtual and augmented reality devices. Association for Computing Machinery, Inc. ISBN 9781450358989; 2018,doi:10.1145/3229434.3229438.
[21] Gruenefeld, U, Koethe, I, Lange, D, Weis, S, Heuten, W. Comparing techniques for visualizing moving out-of-view objects in head-mounted virtual reality. ISBN 9781728113777; 2019, p. 742–746. doi:10.1109/VR.2019.8797725.
[22] Gugenheimer, J, Wolf, D, Haas, G, Krebs, S, Rukzio, E. Swivrchair: A motorized swivel chair to nudge users' orientation for 360 degree storytelling in virtual reality. Association for Computing Machinery. ISBN 9781450333627; 2016, p. 1996–2000. doi:10.1145/2858036.2858040.
[23] Harada, Y, Ohyama, J. Quantitative evaluation of visual guidance effects for 360-degree directions. Virtual Reality 2022;26:759–770. doi:10.1007/s10055-021-00574-7.
[24] Hu, S, Malloch, J, Reilly, D. A comparative evaluation of techniques for locating out-of-view targets in virtual reality. 2021,.
[25] Jo, H, Hwang, S, Park, H, Ryu, JH. Aroundplot: Focus+context interface for off-screen objects in 3d environments. Computers and Graphics 2011;35:841–853. doi:10.1016/j.cag.2011.04.005.
[26] Lin, YC, Chang, YJ, Hu, HN, Cheng, HT, Huang, CW, Sun, M. Tell me where to look: Investigating ways for assisting focus in 360° video. vol. 2017-May. Association for Computing Machinery. ISBN 9781450346559; 2017, p. 2535–2545. doi:10.1145/3025453.3025757.
[27] Markov-Vetter, D, Luboschik, M, Islam, AT, Gauger, P, Staadt, O. The effect of spatial reference on visual attention and workload during viewpoint guidance in augmented reality. Association for Computing Machinery, Inc. ISBN 9781450379434; 2020,doi:10.1145/3385959.3418449.
[28] Renner, P, Pfeiffer, T. Augmented reality assistance in the central field-of-view outperforms peripheral displays for order picking: Results from a virtual reality simulation study. Institute of Electrical and Electronics Engineers Inc. ISBN 9780769563275; 2017, p. 176–181. doi:10.1109/ISMAR-Adjunct.2017.59.
[29] Feiner, S, Macintyre, B, Seligmann, D. Knowledge-based augmented reality. Commun ACM 1993;36(7):53–62. URL: https://doi.org/10.1145/159544.159587. doi:10.1145/159544.159587.
[30] Petersen, N, Pagani, A, Stricker, D. Real-time modeling and tracking manual workflows from first-person vision. In: 2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR). 2013, p. 117–124. doi:10.1109/ISMAR.2013.6671771.
[31] Khuong, BM, Kiyokawa, K, Miller, A, La Viola, JJ, Mashita, T, Takemura, H. The effectiveness of an ar-based context-aware assembly support system in object assembly. In: 2014 IEEE Virtual Reality (VR). 2014, p. 57–62. doi:10.1109/VR.2014.6802051.
[32] Grogorick, S, Stengel, M, Eisemann, E, Magnor, M. Subtle gaze guidance for immersive environments. Association for Computing Machinery, Inc. ISBN 9781450351485; 2017,doi:10.1145/3119881.3119890.
[33] Danieau, F, Guillo, A, Dore, R. Attention guidance for immersive video content in head-mounted displays. 2017, p. 205–206. doi:10.1109/VR.2017.7892248.
[34] Renner, P, Blattgerste, J, Pfeiffer, T. A path-based attention guiding technique for assembly environments with target occlusions. In: 2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR). 2018, p. 671–672. doi:10.1109/VR.2018.8446127.
[35] Borst, CW, Baiyya, VB, Best, CM, Kinsland, GL. Volumetric windows: Application to interpretation of scientific data, shader-based rendering method, and performance evaluation. In: Proceedings of the 2007 International Conference on Computer Graphics & Virtual Reality, CGVR. CSREA Press; 2007, p. 72–80.
[36] Holm, S. A simple sequentially rejective multiple test procedure. Scandinavian Journal of Statistics 1979;6(2):65–70. URL: http://www.jstor.org/stable/4615733.
[37] Healey, C, Enns, J. Attention and visual memory in visualization and computer graphics. IEEE Transactions on Visualization and Computer Graphics 2012;18:1170–1188. doi:10.1109/TVCG.2011.127.

Appendix A. Preliminary Tuning Study

Appendix A.1. Overview and Environment

A preliminary study was conducted to guide the selection of good cue parameters for fairer comparison of cues in the main study (selecting reasonable and consistent parameters). We used an 8x2 within-subjects design. Independent variables were the 8 visual cues and 2 target spacing options (narrow and wide layouts). Highlight was excluded.

All experiments used a virtual offshore oil rig environment, with users seated at a station near the top of the rig. Blue barrels and other environment objects (Figure 2) were used as targets. For narrow spacing, there were 4 target barrels, with moderate space between them, all visible at once, and placed on a platform. For wide spacing, there were 2 barrels spread out on the platform and a helicopter and crane as added targets to the left and right, not fitting in a single view.

Appendix A.2. Participants and Apparatus

13 subjects participated in the tuning study. Each subject used a Vive Pro Eye headset with one controller used for ray-based menu interaction. Experiments were run on an Alienware Aurora R9 with an Intel i7-9700K processor, 64GB of RAM, and GeForce RTX 2080 Super graphics card.

Appendix A.3. Parameters

The cue parameters for each cue are shown in Table A.8. In terms of visuals, most cues share two common parameters: the rate at which the cue fades when gaze approaches the target and cue size. Those with a placement parameter are either placed along an arc (the shortest path along a head-centered sphere) or cubic Hermite curve. In order to handle objects in the VR scene occluding our cues, select cues are rendered such that they appear through other objects.

Table A.8. Parameters for each Visual Cue

Cue	Parameter 1	Parameter 2	Parameter 3
Highlight	N/A	N/A	N/A
Arrow	Offset From Gaze	Fade	Arrow Size
Arrow Trail	Placement	Fade	Arrow Size
Arrow Field	Number of Arrows	Fade	Arrow Size
DynSWAVE	Wave Speed	Fade	Intensity
Border	Boundary	Fade	Movement
Vignette	Darkness	Close Speed	Aperture Size
Line Strip	Placement	Fade	Line Width
Attention Funnel	Rings or Goals	Fade	Ring/Goal Size

Table A.9. The most often-selected parameters for each cue in the tuning study. A / indicates a tie between 2 values.

Arrow	**Offset**	**Fade**	**Size**
Narrow	45%	Small	Medium
Wide	15%	Small	Medium/Large
Arrow Trail	**Placement**	**Fade**	**Size**
Narrow	Along Arc	Small	Medium
Wide	Along Arc	Small	Medium
At Funnel	**Rendering**	**Fade**	**Size**
Narrow	Rings	Small	Small
Wide	Goalpost Edges	Medium	Small
Arrow Field	**# of Arrows**	**Fade**	**Size**
Narrow	Medium	Small	Medium
Coarse	Medium	Small	Medium
Line Strip	**Placement**	**Fade**	**Size**
Narrow	Along Curve	Small	Medium
Wide	Along Curve	Small	Medium
Vignette	**Close Speed**	**Darkness**	**Size**
Narrow	Medium	Dark	Small
Wide	Medium	Light	Medium
Border	**Boundary**	**Fade**	**Movement**
Narrow	Narrow/Medium	Small	Ping Pong
Wide	Medium	Small/Medium	Standard
DynSWAVE	**Intensity**	**Fade**	**Wave Speed**
Narrow	High	Small	Medium
Wide	High	Medium	Medium

Appendix A.4. Procedure

After filling out a consent form, a background questionnaire, and undergoing eye tracking calibration, each subject performed two stages of tuning (one each for wide/narrow layouts). Each stage presented the 8 visual cues in a random order. Subjects changed the three parameters of each cue to find values that they believed to best guide their attention to the target. Initial parameter values were randomized per cue.

Subjects were asked to "adjust all settings to best draw your attention to the correct blue barrel when you are looking at the nearby blue barrels. But avoid settings you find too uncomfortable" (slightly reworded for the wide spacing). In addition to tuning parameter values, subjects were asked which cues they like and disliked, and they were asked to explain (this aspect is not detailed here because it is studied in depth in the main experiment).

In the narrow target spacing stage, participants were placed before a group of four closely spaced blue barrels. In the wide target spacing stage, the number of barrels was reduced to 2 and a helicopter and crane were added as targets, with target spacing placing some targets out of the field of view at all times.

Appendix A.5. Results

Table A.9 shows the most commonly selected parameter values for each cue with wide and narrow targets.

For Arrow, the main difference in parameter choice between narrow and wide targets was the offset. The most commonly picked value for offset with narrow targets was 45%, whereas it was only 15% with wide targets. This is likely due to the fact that when viewing similar objects in a small area, the lower offset clears up any ambiguity. We additionally see a tie between arrow size selections for wide targets. This may indicate that some subjects prefer a larger Arrow when looking for objects spread throughout a scene.

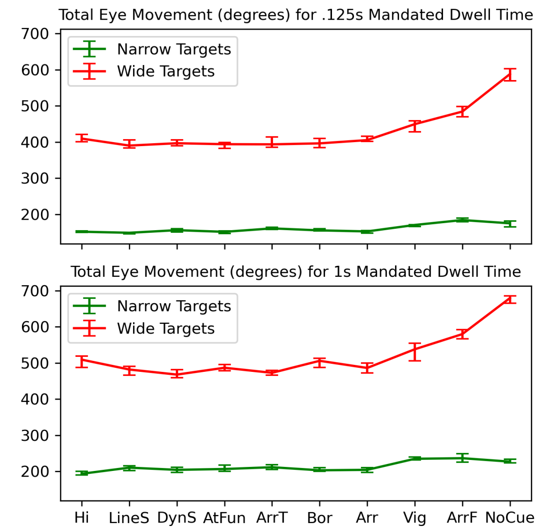

Fig. A.10. Amount of eye movement performed throughout an 8-target trial for each cue (in degrees).

Differing from Arrow, participants most commonly selected small sizes for the Attention Funnel objects. This may be due to its obtrusiveness, which one subject noted. In terms of the rendering technique used, subjects preferred the rings for fine tuning and the goalpost edges for coarse tuning. For consistency, we used the rings rendering technique in the main comparison study.

For Vignette, the majority of subjects chose a dark, small tunnel for narrow targets and a light, medium sized tunnel for wide targets. Based on this, it appears that a more aggressive tunnel may be necessary to find the correct object with narrow targets, where the target object may be hard to discern.

There were no differences in the most commonly selected parameter values for narrow and wide targets with the Arrow Field. Subjects preferred the medium setting for # of arrows, which is not in line with subject feedback in later studies, where subjects indicate there are not enough arrows for narrow targets. This may be a problem inherent to the cue, though, as one subject noted that the cue is ambiguous regardless of the setting.

The most common DynSWAVE selections were similar between narrow and wide targets, with the only difference being the fade rate. Despite saying that the cue was disorienting, participants chose a high intensity (many smaller waves) and medium wave speed.

There were no differences in selected settings between narrow and wide targets for either Line Strip or Arrow Trail. Both cues displayed a continuous object(s) along an arc or curve, but subjects chose the arc version for Arrow Trail and the curve version for Line Strip.

For Border, only the selected movement setting differed be-

Table A.10. Pairwise comparisons for total eye movements between different cues. One pairwise test was performed for each combination of Target Spacing and Dwell Time levels, making a total of 4. Number and color show the number of significant differences found between the pairs ($p < .05$ after Holm correction). Sign indicates if the row's cue had a lower total amount of movement than the column cue.

Cue	Hi	LineS	DynS	AtFun	ArrT	Bor	Arr	Vig	ArrF	NoCue
Hi	0	0	0	0	0	0	0	-3	-4	-4
LineS	0	0	0	0	0	0	0	-4	-4	-3
DynS	0	0	0	0	0	0	0	-4	-4	-3
AtFun	0	0	0	0	0	0	0	-4	-4	-3
ArrT	0	0	0	0	0	0	0	-4	-4	-3
Bor	0	0	0	0	0	0	0	-3	-4	-3
Arr	0	0	0	0	0	0	0	-4	-4	-4
Vig	3	4	4	4	4	3	4	0	-1	-2
ArrF	4	4	4	4	4	4	4	1	0	-2
NoCue	4	3	3	3	3	3	4	2	2	0

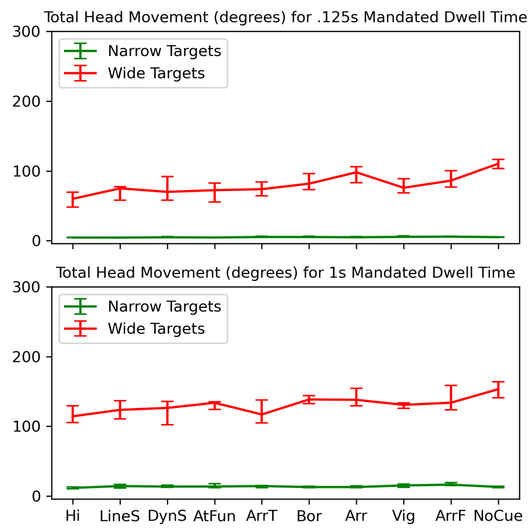

Fig. B.11. Amount of head movement performed throughout an 8-target trial for each cue (in degrees).

tween wide and narrow targets. This may be partly due to confusion about how Border's movement parameter worked, as the ping pong effect is minimally noticeable with high amounts of eye movement.

When carrying over tuned settings to restoration and guidance stages, we decided to keep the fade parameter consistent across cues and target spacings to remove fade as a confounder. Given that a small fade was chosen for most (11) conditions, we set all fade parameters to small. To ensure subjects would not view rings and goalpost versions of Attention Funnel as two separate cues, we set rendering method to rings for all conditions. When two settings tied (e.g. medium and large arrows with wide targets), we chose the setting that made the cue consistent between target spacings (e.g. medium for arrow size).

Appendix B. Additional Metrics

Other secondary measures were taken during the attention guidance stage. After analysis, these were considered to not substantially augment the current findings, but are included here for completeness. Other secondary measures include total eye and head movements. These movements are summed from the changes in angle between gaze or head-forward vectors and vectors to the target on every frame.

Table B.11. Pairwise comparisons for total head movements between different cues. One pairwise test was performed for each combination of Target Spacing and Dwell Time levels, making a total of 4. Number and color show the number of significant differences found between the pairs ($p < .05$ after Holm correction). Sign indicates if the row's cue had a lower (better) time than the column cue.

Cue	Hi	LineS	DynS	AtFun	ArrT	Bor	Arr	Vig	ArrF	NoCue
Hi	0	0	0	0	0	0	-2	0	-4	-2
LineS	0	0	0	0	0	0	-2	0	-3	-2
DynS	0	0	0	0	0	0	-1	0	-2	-2
AtFun	0	0	0	0	0	0	-1	0	-2	-2
ArrT	0	0	0	0	0	0	-2	0	-2	-2
Bor	0	0	0	0	0	0	1	0	0	-2
Arr	2	2	1	1	2	-1	0	1	0	-2
Vig	0	0	0	0	0	0	-1	0	-1	-2
ArrF	4	3	2	2	2	0	0	1	0	-1
NoCue	2	2	2	2	2	2	2	2	1	0

Subject eye and head movements were captured throughout each guidance task trial. Movements were totaled from per-frame angular (degree) changes throughout the trial, and are summarized in Figures B.11 and A.10. Movement totals follow a similar distribution as time metrics, and as such are analyzed with Friedman and pairwise Wilcoxon tests. Analyses are summarized in Tables B.11 and A.10. We assume a better cue will incur lower amounts of movement, as more movement likely increases the amount of time to find the target and, after repeated attempts, may cause fatigue.

Eye movements follow a similar trend as guidance time, though with fewer significant differences between cues. Only Vignette, ArrowField, and NoCue are shown to incur significantly more eye movement than the rest of the cues. Not surprisingly, the wide targets case incurred much more eye movement, and No Cue shows a larger spike in movement than with narrow targets.

Head movements follow a similar pattern as eye movements with a few notable exceptions. Head movement is near 0 for all cues in the narrow targets case. Only Arrow Field differs from any other cue when targets are narrow, likely implying that subjects had to move their heads to see the various arrows around the scene. When targets were wide, Arrow incurred significantly more head movement than most other cues. We expect this is due to the arrow's offset between the gaze point and the target; it may have encouraged users to "chase" the cue with head movements in addition to eyes.

Interestingly, Vignette was shown to incur more eye movement than most cues (in addition to performing worse in most other metrics), but did not differ significantly from most cues in head movement. We infer that something about the cue's design encouraged subjects to search through targets with their eyes, but maintain minimal motion with their heads.