Design and Evaluation of Visual Cues for Restoring and Guiding Visual Attention in Eye-Tracked VR

Jason W. Woodworth Andrew Yoshimura Nicholas G. Lipari Christoph W. Borst

CACS VR Lab
University of Louisiana at Lafayette

Abstract

Distraction can be a problem in VR training environments. We investigate 9 visual cues intended to guide or restore attention to objects of interest to mitigate distraction. A survey of related literature suggests a past focus on "search and selection" tasks to evaluate a cue's capability for guidance. We compare this to a new type of task that focuses on how to restore attention when a short distraction (e.g. a notification) shifts focus away from an object. 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. Factors such as object spacing, gaze dwell time, and distraction distance and duration are varied. The results 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. One implication for more complex distraction tasks considers 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. In addition to revealing the differences between guidance and restoration performance of cues, substantial depth is added beyond prior work by the broader range of conditions and cues included.

Index Terms: Human-centered computing—Human computer interaction (HCI)—HCI design and evaluation methods—User studies; Human-centered computing—Visualization—Visualization design and evaluation methods

1 Introduction

Mixed reality is quickly becoming a key component in the transition to Industry 4.0. Virtual Reality, in particular, is often deployed as a training solution for minimizing cost and human risk [23]. Attention loss is a common concern in such training applications. A user, for example, may miss important details if not looking at a target object while it is being explained. Visual cues, such as arrows [2, 4, 13] or highlights [20, 24] on target objects, are often employed to guide a user's attention to relevant objects, and as such have been the subject of substantial research. Less work, however, has focused on how to guide a user's attention once it has already been lost.

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 user attention shifts elsewhere (attention restoration) or to 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.

For guidance, users are visually guided by cues to a set of targets in a random sequence. In this we measure total time to get through a whole set. For restoration, this sequence is occasionally interrupted by distraction events, after which gaze is restored by a visual cue, measuring 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 vary 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 elements (1, 3, or 5).

Through our study results, we find that a highlight, DynSWAVE [27], and a thick line strip rendered along a Hermite curve are generally the best performing cues, though each perform 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 these problems 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) deep insight into how each cue performs in varying conditions (wide vs narrow target layouts, different distraction parameters) and recommendations for particular training contexts.

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 our surveyed works, and is often used as a baseline to test a new cue against (e.g. [4, 11, 22]). Arrows typically perform well [22, 27], but not exceptionally so, and may have difficulty guiding to targets behind the user [15]. Arrows are most often rendered as 3D objects pointing at the object's position in the scene [4, 18] or a 2D image rendered pointing in the direction the user should look to face the target [21, 27]. The FlyingARrow [13, 17], 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 form of simple visual overlay or outline of the target object [7, 19, 25]. 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 [1, 9]. Other works insert highlights diegetically (e.g. a firefly [24] or swarm of bugs [20]) to draw attention while maintaining immersion.

A common limitation with highlighting techniques is the require-

This is an author-formatted version. Original publication: J. W. Woodworth, A. Yoshimura, N. G. Lipari and C. W. Borst, "Design and Evaluation of Visual Cues for Restoring and Guiding Attention in Eye-Tracked VR", 2023 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops (VRW), Shanghai, China, 2023, pp. 442-450, doi: 10.1109/VRW58643.2023.00096.

Author	Cues	Task	Metrics	Results	ET	Subjects	Medium
Bailey, 2009 [1]	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 [2]	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 [3]	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 [4]	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 [8]	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 [11]	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 [10]	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 [13]	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 [12]	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 [14]	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 [15]	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 [17]	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 [18]	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 [20]	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 [21]	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 [22]	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 [27]	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.

ment 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. [9] 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 [4, 8, 10], but are sometimes just placed in front [11, 20, 22] 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 seen 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 our surveyed papers did not include eye-tracking within their cues' designs. That is, the appearance of the cue responds to the user's eye movements instead of being placed statically in a scene or fixed to head gaze. Studies that included eye-tracking suggest that it has potential to improve visual cues [27]; 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).

3 Visual Cues Design

Figure 1 shows our visual cues in the test environment with a narrow target spacing. We tuned the cues to reasonable appearance parameters within a separate pilot study. 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 a path to the target. As with Arrow, it is placed along the shortest arc path. As the gaze moves away from the target, the number of arrows increases. We expect his 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 point 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 [27]. 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 [27, 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 the boundary. This cue is similar to Halo [10], however, considering our experiment environment, we do not adjust for clutter.

Vignette (Vig) Our *Vignette* (Figure 1f) cue is similar to Fade to Black [6]; 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 [10] and 3DPath [26]. A thick line strip (appearing like a curved banner) is displayed along a 3D Hermite cubic curve, showing the full path from a point on the gaze vector (at a small distance from the head) to the target. 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. [3], 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.

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 [5]). All cues fade according to gaze such that they appear after looking 3 degrees away from the target and become more apparent as the user looks away.

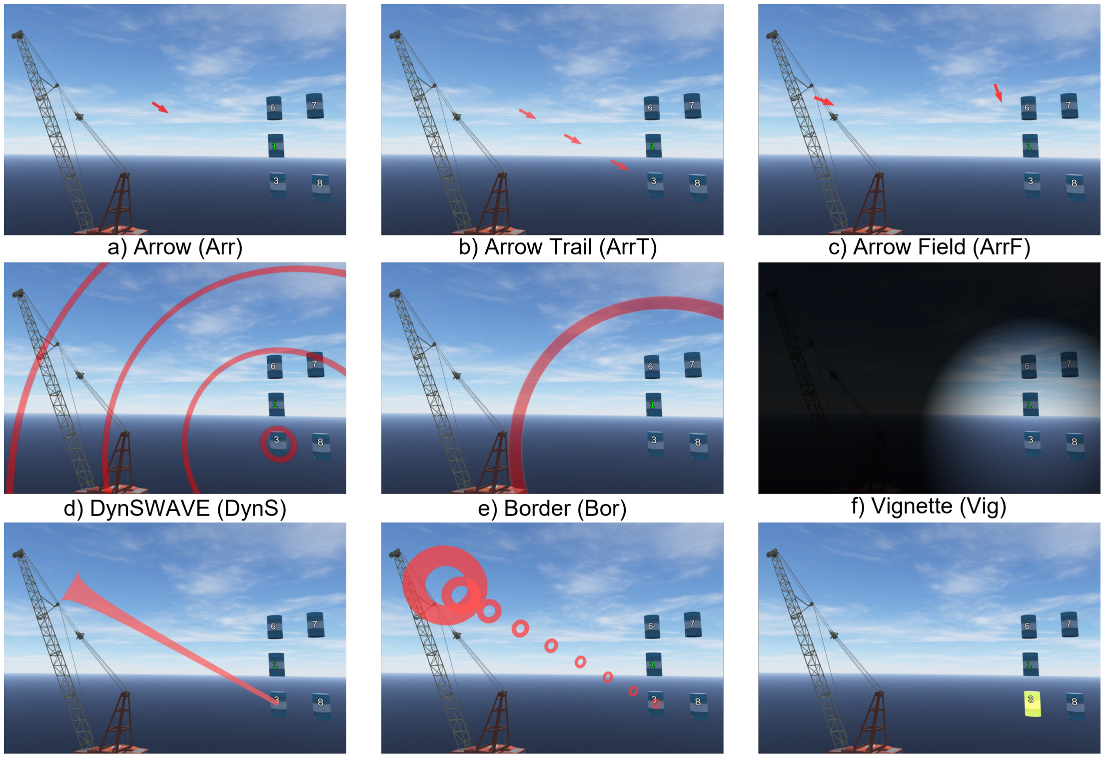
4 Guidance and Restoration Studies

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 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, 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



a) Arrow (Arr) b) Arrow Trail (ArrT) c) Arrow Field (ArrF)
d) DynSWAVE (DynS) e) Border (Bor) f) Vignette (Vig)
g) Line Strip (LineS) h) At Funnel (AtFun) i) Highlight (Hi)

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

R13 with an Intel 17-12700KF processor and GeForce RTX 3080 graphics card.

4.2 Barrel Setup and Patterns

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

Attention Restoration Stage The 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 is 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, for 1 second each. While gaze dwells on a barrel, its label turns light green. Once the target barrel has been gazed at for 1 second, it is "activated" and its label turns a darker shade of green. For every 2 or 3 activated barrels (random), a distraction appears while the subject dwells on a barrel. This happens after 0.5-1 seconds of dwell (random). Once the distraction task begins, a visual cue is rendered to help guide them back. After activating a full set of 8 barrels, the barrel pattern is re-randomized and reset. This is done as many times as needed until 6 distraction events have occurred.

To handle the distraction objects, a subject must look at each sphere, in an order, removing the spheres. Distractions appear with a "beep" sound. Spheres are placed on an arc such that they have equal distance from the target barrel. A sphere at one end of the arc (chosen randomly) is red and others (if present) are initially white. The subject must look at the red sphere, causing it to vanish and the next sphere along the arc to turn red. Each red sphere disappears when it is gazed at for 3 graphics frames. The 3-frame threshold avoids activation by jitter or approximate moves over the distraction.

Distraction Distance has 2 levels: close and far. Close and far

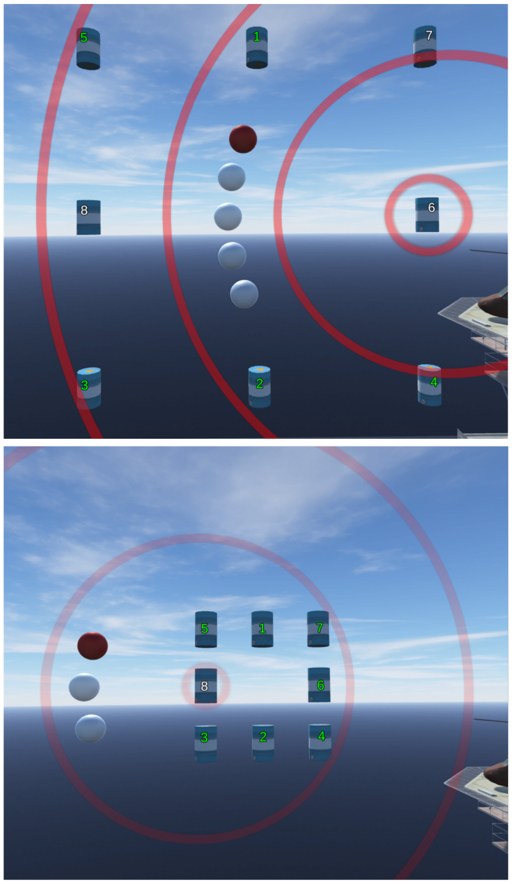


Figure 2: Environment layouts used in experiments. Top: wide target spacing with 5 far distractions during restoration stage. Bottom: narrow target spacing with 3 close distractions during restoration stage.

spheres have a distance of approximately 4.2 and 7.1 meters from the target barrel, respectively. Distraction Breadth has 3 levels: 1, 3, or 5 distraction spheres. For 1 sphere, the sphere is positioned either vertically/horizontally (for the middle barrels) or diagonally (for corner barrels) away from the barrels by the sphere distance. When Target Spacing is narrow, distractions appear outside of the barrel layout; when it is wide, distractions appear within the barrel layout (as seen in Figure 2).

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

Guidance Stage The guidance stage used a 10x2x2 study design with the following 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.

Subjects were given the following instructions with in-VR text

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

Table 2: Pairwise 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). Sign indicates if the row's cue had a lower (better) time than the column cue.

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."

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.

4.4 Results

4.4.1 Metrics

We consider cues that can quickly guide or restore attention to be superior. To that end, we use time as our primary metric. 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. For the guidance stage, the **guidance 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.

Much of the data shows positive skew, with high-time outliers. So, analysis primarily uses Friedman tests with Wilcoxon signed-rank pairwise followups. Due to the large number of pairs for 10 visual cues, we correct p values using Holm corrections [16] to avoid familywise error inflation. While multiple independent variables are used in study, we focus primarily on differences between cues.

4.4.2 Guidance Condition Time

Total guidance times between cues with all levels of other variables included are summarized in Figure 3. A Friedman test showed differences between these cue conditions ($\chi^2(9) = 472.15, p < .001$). 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

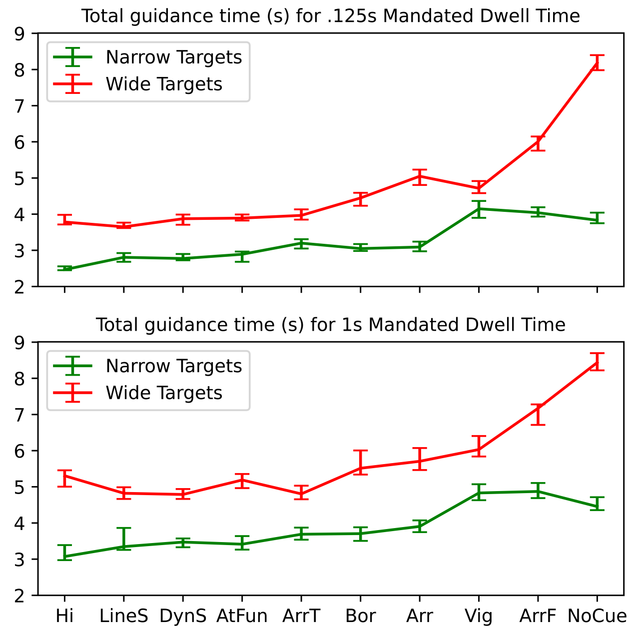


Figure 3: Median guidance time for each cue at different levels of Target Spacing and Dwell Time. Error bars were computed by bootstrapping and show the range containing the middle 68 percent of 5000 median estimates, each being the median of 98 values (sampled with replacement). This is analogous to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

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

Table 3: Pairwise comparisons between cues for restoration lookback time. A negative (-) sign indicates less lookback time for the row's cue.

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.

4.4.3 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 4. The most notable result is that median times for the cues do not follow the same patterns as they did in the guidance phase. 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 4 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.

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.

5.1 Visual Cues Discussion

5.1.1 Highlight

Highlight is one of the best-performing cues. For attention restoration (Table 3), 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. We suspect the quick change in color naturally draws the user's eye; subjects noted the effect occasionally making their eyes move to it automatically before the were conciously aware of it.

Important to Highlight's design is that the effect is placed on the target itself. A potential benefit to this that it draws attention to the target directly rather than indirectly through another visual effect. A possible drawback is if the user does not have the target in view, there should be no effect on them. Thus, we expected Highlight to perform poorly with wide Target Spacing. However, the difference between narrow and wide does not appear much larger than 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 1.5 seconds longer to complete. We expect subjects did some broader search before catching the Highlight effect in their periphery, then moved quickly towards the target. More insight could be found by further breaking down how time was spent during the guidance task.

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 better performance than many cues (Table 2). The design

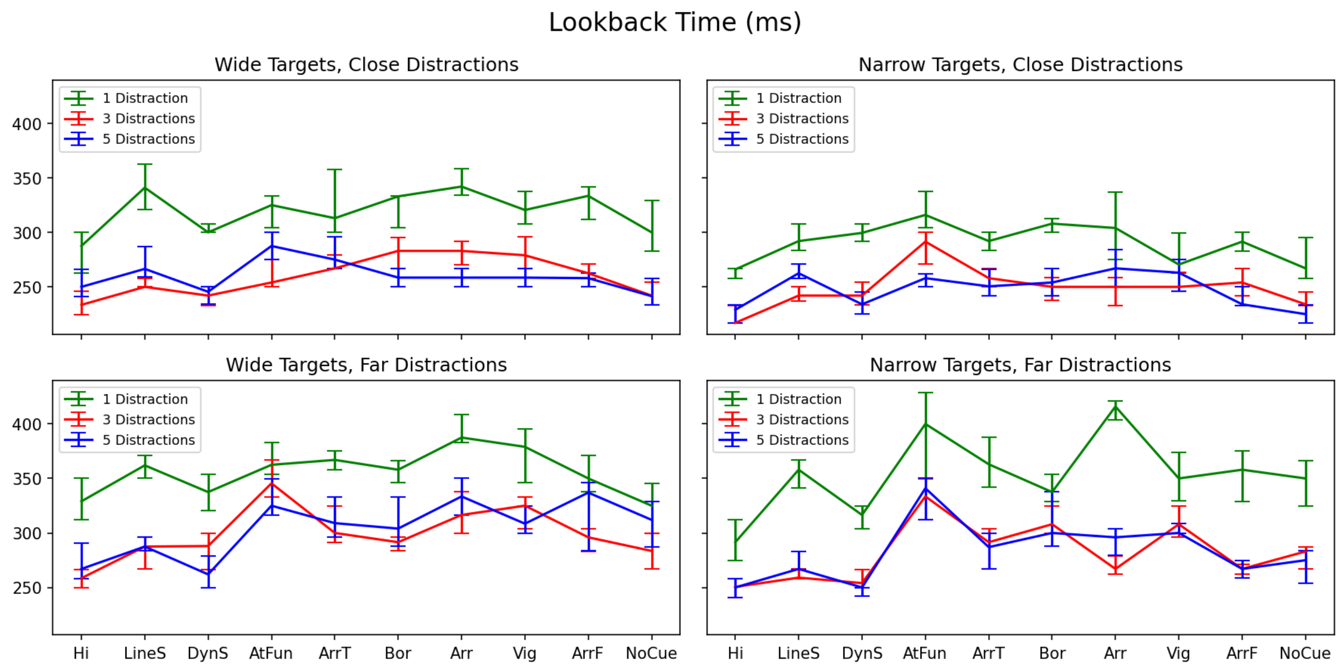


Figure 4: Median time to look back from distractions to correct target for each cue at different levels of Distraction Breadth, Distraction Distance, and Target Spacing. Error bars were computed by bootstrapping and show the range containing the middle 68 percent of 5000 median estimates, each being the median of 67 values (sampled with replacement).

of the cue makes it immediately noticeable, which likely reduces delay to move to the next target. We expected a longer search time, given that cues with animated speeds have sometimes been shown to induce eye movement at that speed [17], but this does not appear to be the case here.

DynSWAVE outperforms Border in several cases while Border never performs better than DynSWAVE (Table 2). According to subjects' opinions, 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) and makes it impractical for extending to multiple targets. 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 will render 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 4), 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 3). However, both perform much better in the guidance task, showing better times than 6 of the other cues in at least 1 condition (Table 2), and score much closer to each other.

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 right in front of the user appears to be useful in guidance tasks, it may actually be a hindrance and distraction in restoration tasks.

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 3), 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 during an initial parameter tuning pilot, 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 taks (Figure 4), 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.

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. 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.

5.1.5 Vignette

Vignette has one of the highest median lookback times in the restoration task (Figure 4); it is shown to be worse than Highlight, DynSWAVE, and No Cue. This trend continues in the guidance task, where for total guidance time it is only shown to be better than No Cue and Arrow Field. In subject closing subject interviews, they often said they did not like how obtrusive it was and it was too indirect to see exactly where it was pointing to.

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. This potential positive may have been outweighed by the fact that, like Border, Vignette uses a "circular" style of cue that indirectly suggests a broad potential of target locations instead of a precise targeted area.

5.1.6 Arrow Field

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

5.1.7 No Cue

No Cue performs surprisingly well in attention restoration, having the third lowest median lookback time and shown to be better than Vignette and Attention Funnel (Table 3). This changes dramatically in the guidance task where it's worse than every other cue in at least one case (Table 2), particularly with 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 in results 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 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 we also see particularly large gaps for No Cue between near and far distractions (Figure 4), suggesting that when the subjects gazed far enough away from the target that it gets out of view, it took longer to find again without assistance. Were there a task with distractions even further (e.g. behind the user) we may see No Cue's performance get even worse.

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 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 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 are 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.

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 obtrustive for the application environment, LineStrip along a curve will provide quick guidance and may be usable with multiple targets, but may not be appropriate for a restoration task. In fact, a designer may opt to show no cue for a restoration task unless they can confirm the user is actually mentally distracted or the distraction is complex.

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 prior work is not sufficient to apply directly to attention restoration. Given the prevalence of distraction in VR environments, more research may need to be done specifically into restoration tasks.

Limitations and Future Work While we accounted for many variables within our restoration and guidance tasks, there is still a variety of conditions cues need to be tested in 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 break down 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] R. Bailey, A. McNamara, N. Sudarsanam, and C. Grimm. Subtle gaze direction. *ACM Transactions on Graphics*, 28:1–14, 8 2009. doi: 10.1145/1559755.1559757

[2] N. Binetti, L. Wu, S. Chen, E. Kruijff, S. Julier, and D. P. Brumby. Using visual and auditory cues to locate out-of-view objects in head-mounted augmented reality. *Displays*, 69, 2021. doi: doi.org/10.1016/j.displa.2021.102032

[3] F. Biocca, A. Tang, C. Owen, and F. Xiao. Attention funnel: Omnidirectional 3d cursor for mobile augmented reality platforms. pp. 1115–1122, 2006.

[4] F. Bork, C. Schnelzer, U. Eck, and N. Navab. Towards efficient visual guidance in limited field-of-view head-mounted displays. *IEEE Transactions on Visualization and Computer Graphics*, 24:2983–2992, 11 2018. doi: 10.1109/TVCG.2018.2868584

[5] C. W. Borst, V. B. Baiyya, C. M. Best, and G. L. Kinsland. 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*, pp. 72–80. CSREA Press, 2007.

[6] F. Danieau, A. Guillo, and R. Dore. Attention guidance for immersive video content in head-mounted displays. pp. 205–206, 01 2017. doi: 10.1109/VR.2017.7892248

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

[8] S. Grogorick, G. Albuquerque, and M. Magnor. Comparing unobtrusive gaze guiding stimuli in head-mounted displays. pp. 2805–2809, 2018.

[9] S. Grogorick, M. Stengel, E. Eisemann, and M. Magnor. Subtle gaze guidance for immersive environments. Association for Computing Machinery, Inc, 9 2017. doi: 10.1145/3119881.3119890

[10] U. Gruenefeld, A. E. Ali, S. Boll, and W. Heuten. Beyond halo and wedge: Visualizing out-of-view objects on head-mounted virtual and augmented reality devices. Association for Computing Machinery, Inc, 9 2018. doi: 10.1145/3229434.3229438

[11] U. Gruenefeld, D. Ennenga, A. E. Ali, W. Heuten, and S. Boll. Eyesee360: Designing a visualization technique for out-of-view objects in head-mounted augmented reality. pp. 109–118. Association for Computing Machinery, Inc, 10 2017. doi: 10.1145/3131277.3132175

[12] U. Gruenefeld, I. Koethe, D. Lange, S. Weis, and W. Heuten. Comparing techniques for visualizing moving out-of-view objects in head-mounted virtual reality. pp. 742–746, 2019. doi: 10.1109/VR.2019.8797725

[13] U. Gruenefeld, D. Lange, L. Hammer, S. Boll, and W. Heuten. Flyingarrow: Pointing towards out-of-view objects on augmented reality devices. Association for Computing Machinery, Inc, 6 2018. doi: 10.1145/3205873.3205881

[14] J. Gugenheimer, D. Wolf, G. Haas, S. Krebs, and E. Rukzio. Swivrchair: A motorized swivel chair to nudge users' orientation for 360 degree storytelling in virtual reality. pp. 1996–2000. Association for Computing Machinery, 5 2016. doi: 10.1145/2858036.2858040

[15] Y. Harada and J. Ohyama. Quantitative evaluation of visual guidance effects for 360-degree directions. *Virtual Reality*, 26:759–770, 6 2022. doi: 10.1007/s10055-021-00574-7

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

[17] S. Hu, J. Malloch, and D. Reilly. A comparative evaluation of techniques for locating out-of-view targets in virtual reality. 2021.

[18] H. Jo, S. Hwang, H. Park, and J. H. Ryu. Aroundplot: Focus+context interface for off-screen objects in 3d environments. *Computers and Graphics*, 35:841–853, 2011. doi: 10.1016/j.cag.2011.04.005

[19] B. M. Khuong, K. Kiyokawa, A. Miller, J. J. La Viola, T. Mashita, and H. Takemura. The effectiveness of an ar-based context-aware assembly support system in object assembly. In *2014 IEEE Virtual Reality (VR)*, pp. 57–62, 2014. doi: 10.1109/VR.2014.6802051

[20] D. Lange, T. C. Stratmann, U. Gruenefeld, and S. Boll. Hivefive: Immersion preserving attention guidance in virtual reality. Association for Computing Machinery, 4 2020. doi: 10.1145/3313831.3376803

[21] Y. C. Lin, Y. J. Chang, H. N. Hu, H. T. Cheng, C. W. Huang, and M. Sun. Tell me where to look: Investigating ways for assisting focus in 360° video. vol. 2017-May, pp. 2535–2545. Association for Computing Machinery, 5 2017. doi: 10.1145/3025453.3025757

[22] D. Markov-Vetter, M. Luboschik, A. T. Islam, P. Gauger, and O. Staadt. The effect of spatial reference on visual attention and workload during viewpoint guidance in augmented reality. Association for Computing Machinery, Inc, 10 2020. doi: 10.1145/3385959.3418449

[23] J. E. Naranjo, D. G. Sanchez, A. Robalino-Lopez, P. Robalino-Lopez, A. Alarcon-Ortiz, and M. V. Garcia. A scoping review on virtual reality-based industrial training. *Applied Sciences*, 10(22), 2020. doi: 10.3390/app10228224

[24] L. T. Nielsen, M. B. Møller, S. D. Hartmeyer, T. C. Ljung, N. C. Nilsson, R. Nordahl, and S. Serafiny. Missing the point: An exploration of how to guide users' attention during cinematic virtual reality. vol. 02-04-November-2016, pp. 229–232. Association for Computing Machinery, 11 2016. doi: 10.1145/2993369.2993405

[25] N. Petersen, A. Pagani, and D. Stricker. Real-time modeling and tracking manual workflows from first-person vision. In *2013 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pp. 117–124, 2013. doi: 10.1109/ISMAR.2013.6671771

[26] P. Renner, J. Blattgerste, and T. Pfeiffer. A path-based attention guiding technique for assembly environments with target occlusions. In *2018 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pp. 671–672, 2018. doi: 10.1109/VR.2018.8446127

[27] P. Renner and T. Pfeiffer. Attention guiding techniques using peripheral vision and eye tracking for feedback in augmented-reality-based assistance systems. pp. 186–194. Institute of Electrical and Electronics Engineers Inc., 4 2017. doi: 10.1109/3DUI.2017.7893338

[28] P. Renner and T. Pfeiffer. Augmented reality assistance in the central field-of-view outperforms peripheral displays for order picking: Results from a virtual reality simulation study. pp. 176–181. Institute of Electrical and Electronics Engineers Inc., 10 2017. doi: 10.1109/ISMAR-Adjunct.2017.59