Directing versus Attracting Attention: Exploring the Effectiveness of Central and Peripheral Cues in Panoramic Videos

Anastasia Schmitz*   Andrew MacQuarrie   Simon Julier   Nicola Binetti   Anthony Steed

University College London
Department of Computer Science

ABSTRACT

Filmmakers of panoramic videos frequently struggle to guide attention to Regions of Interest (ROIs) due to consumers' freedom to explore. Some researchers hypothesize that peripheral cues *attract* reflexive/involuntary attention whereas cues within central vision engage and *direct* voluntary attention. This mixed-methods study evaluated the effectiveness of using central arrows and peripheral flickers to guide and focus attention in panoramic videos. Twenty-five adults wore a head-mounted display with an eye tracker and were guided to 14 ROIs in two panoramic videos. No significant differences emerged in regard to the number of followed cues, the time taken to reach and observe ROIs, ROI-related memory and user engagement. However, participants' gaze travelled a significantly greater distance toward ROIs within the first 500 ms after flicker-onsets compared to arrow-onsets. Nevertheless, most users preferred the arrow and perceived it as significantly more rewarding than the flicker. The findings imply that traditional attention paradigms are not entirely applicable to panoramic videos, as peripheral cues appear to engage both involuntary and voluntary attention. Theoretical and practical implications as well as limitations are discussed.

Keywords: Cinematic Virtual Reality, 360° video, head-mounted display, guiding attention, memory, eye-tracking.

Index Terms: Human-centered computing—HCI theory, concepts and models; **Human-centered computing—Virtual Reality**; Human-centered computing—Empirical studies in HCI

1 INTRODUCTION

Panoramic videos, also known as Cinematic Virtual Reality (CVR), allow consumers to experience virtual environments via an omnidirectional view by using desktop or mobile applications [15]. Driven by the emergence of affordable head-mounted displays (HMDs), the popularity and consumption of CVR has greatly increased in recent years [44, 45] and has encouraged its introduction into entertainment [71], education [29] and vocational training [32, 60]. Two key attractions of HMDs are their large Field of View (FOV), which immerses users in virtual environments [18], and their capability to equip users with the freedom to look anywhere in the scene using head movements [8, 15]. The majority of CVR experiences, which are this paper's primary focus, are monoscopic, have fixed-viewpoints, and are consumed passively [45].

In contrast to traditional film media, filmmakers of CVR cannot predict where users will focus their attention and frequently struggle to convey content effectively [7]. Equally, some CVR-users are concerned about missing plot elements due to a lack of guidance [45,67]. In response to this, researchers have suggested that visual cues can be an effective way to guide users to Regions of Interest (ROIs) [66]. Popular techniques include stimulus-based luminance modulations

---

*e-mail: anastasia.schmitz.18@alumni.ucl.ac.uk

(i.e. flickers) in the user's periphery [65], arrows or saliency adjustments to ROIs (i.e. modulations of contrast, saturation and/or color of a particular video region to make it more/less conspicuous) [68]. Few studies have investigated whether these cues encourage users to focus on ROIs and to encode related information into memory [65]. Indeed, attention models and taxonomies, which are largely based on research with desktop computers, suggest that only directional cues within users' central vision enhance attention for sustained amounts of time [31, 35, 66].

As the generalizability of the above assumption to CVR has not yet been investigated, this study explored the effectiveness of a central arrow and a peripheral flicker cue in guiding as well as focusing participants' attention in two multi-ROI panoramic videos. This paper has four contributions. First, it provides guidelines for designing arrows and flickers for CVR. Second, it proposes a novel combination of research methods to investigate user experiences with attention-guiding techniques in CVR. Third, it provides insights into consumers' experience and engagement with arrows and flickers as well as their memory of cued ROIs. Finally, it outlines design implications to aid filmmakers of panoramic videos in guiding users.

2 RELATED WORK

2.1 The Different Types of Panoramic Video

Panoramic videos are created with omnidirectional or multiple traditional cameras and are commonly stitched into equirectangular projections, which wrap around a video sphere with the user at the center [15]. The videos feature content from multiple genres, ranging from action and comedy to horror and news [45]. Although they resemble traditional film media, the composition of a narrative in CVR is different and requires careful consideration of ROIs' time and place in order to form a coherent storyline. Lin and colleagues [43] suggest that there are five different types of storytelling in CVR based on ROI number, location and dynamism in the scene. Guiding and focusing attention is particularly challenging if ROIs are outside of the user's FOV and if videos include "teleportation" as well as multiple dynamic ROIs.

2.2 Subtle versus Overt Cues

Filmmakers of multi-ROI panoramic videos wish for consumers to attend to ROIs in a desired sequence and to look at a particular location at a particular time. Conventional methods, such as close-ups, camera angles and zoom, are inappropriate for CVR due to its wide projection and immersive qualities [7]. Therefore, two methods have been proposed: subtle and overt cues [26].

Subtle cues are designed to be as covert or implicit as possible to avoid adverse effects on the viewing experience. For example, they can take the form of faint peripheral flickers and colour changes [26]. Other examples are character movements, spotlights [39], spatial audio [18] as well as gestures and gaze [7, 39], which are incorporated into storylines upon shooting. Although subtle cues have encouraged participants in previous research to look toward certain regions in 360-degree spaces more often than conditions with no cues [7], researchers cannot guarantee that users will follow them or know that they are expected to do so [26]. In contrast,

overt cues are typically added to panoramic videos post-production. While decreasing users' level of immersion, they appear effective in guiding and focusing users' gaze on ROIs [18]. Indeed, Rothe and colleagues [65] demonstrated that a noticeable peripheral flicker resulted in significantly more gaze shifts toward three static ROIs in a panoramic video compared to a faint flicker. Further examples of overt cues are blurring or fading of less important regions [16, 30], saliency adjustments [68], arrows [43, 62], picture-previews, centrally/peripherally located halos and triangles leading to out-of-view ROIs [27] as well as automatic rotations [42, 43] and distracting objects (e.g. fireflies) [9].

2.3 Reacting to Central versus Peripheral Cues

Researchers have suggested that visual overt cues, which are this paper's focus, differ in effectiveness for guiding attention based on their position within the user's FOV [66]. Posner [59] famously described attention as a spotlight, as it is narrow and directed at fixated objects in the center of gaze. Similarly, visual processing is constrained by the eye's fovea, which is responsible for approximately the central 5-degrees of vision [49] and for perceiving high visual acuity, shapes and colors [23]. These attentional and anatomical constraints engender eye movement strategies required to sample information beyond the currently fixated location of visual space. For example, as the eye's periphery is highly sensitive to light, flickering can attract the "spotlight of attention" toward a ROI [20].

Therefore, cues within central versus peripheral vision typically have different features. Central cues (e.g. arrows) are typically symbolic and ***direct*** users toward ROIs [31, 66]. They explicitly suggest that something outside of the user's focus could be important, attributing significance to an ROI. However, the meaning (e.g. directionality) of central cues needs to be interpreted before users can follow them. Thus, theories of attention suggest that central cues engage voluntary/endogenous attention and can be ignored if their presence is undesired or does not match the individual's intentions [37, 51]. In contrast, peripheral cues (e.g. flickers) ***attract*** attention to ROIs [31, 37] and involve involuntary/exogenous attention. They are harder to ignore and faster to respond to than central cues, as their information is not processed consciously [56, 59, 63].

Psychology experiments have supported the above theories using desktop computers and the Posner cueing task, which assesses participants on their attentional shifts to targets at cued and uncued locations. Studies have shown that participants typically react faster to targets at cued compared to uncued locations, which is termed the "facilitation effect" [57]. In previous research, this effect emerged as soon as targets appeared 50-100 ms after a peripheral flash even if it was non-predictive of the target location [40, 57, 59]. However, for central arrows, it developed only if the target appeared 300 ms or more post-cue [40, 57, 59]. Participants also processed sensory information of a peripheral flash within 150 ms after cue-onset when guided to letter stimuli in their periphery, whereas they experienced a 100 ms processing-delay when presented with central arrows [17]. The findings demonstrate a reflexive response to peripheral cues and conscious/voluntary attention to central cues.

Despite appearing slower than peripheral cues in guiding users to ROIs, central cues have encouraged participants to attend to ROIs for more sustained periods of time in desktop-based research [55]. In one study, involuntary attention only lasted for 300 ms after a peripheral cue-onset whereas voluntary attention lasted for at least 1200 ms [52]. Voluntary attention also increased individuals' vigilance [34]. In contrast, participants experienced an inhibition of attention to peripherally cued ROIs due to lacking the intention to focus on them [1, 35]. Therefore, peripheral cues may be less effective in encouraging users to sustain attention on ROIs.

Little research has assessed if the above traditional attention paradigms generalize to CVR, which can feature out-of-view ROIs that could complicate attention guidance. Grogorick and colleagues

[25] compared several subtle peripheral cues, such as flicker, red dot, and zoom as well as blur techniques, in panoramic images within an immersive dome projection. Approximately 20-40% of participants responded with saccades (i.e. fast eye movements) toward ROIs within the first second after cue-onset for all methods and the luminance modulation was most effective. The researchers, thus, concluded that peripheral cues elicited reflexive responses. Participants' relatively low response rate was attributed to the subtle nature of the cues. Furthermore, Renner and Pfeiffer [62] asked participants to assemble a birdhouse in a 6DoF (i.e. six degrees of freedom) virtual reality experience. Participants were directed toward necessary pieces via several central cueing methods (e.g. an arrow, an image of a tunnel of gates, or an image of waves moving toward the ROI). Although the birdhouse was assembled fastest when the arrow was present, participants had to interpret the arrow's direction before following it, thus, using voluntary attention.

2.4 Memory Effects

Memory is often considered a higher-level cognition than attention because it occurs in the later stages of information processing [20, 31]. Nevertheless, both mechanisms are interrelated. Indeed, peripheral and central cues generally speed information processing at ROIs and slow processing at uncued regions [11, 22]. However, central cues appear to encourage users to process ROI-related information more actively and deeply into explicit memory than peripheral cues [40]. Explicit memory is a type of long-term memory and stores information for conscious recall and recognition of factual information [20]. For example, Hauer and MacLeod [31] showed that participants recognized more words on a desktop screen when they were previously directed to them by a row of central arrows compared to when they were attracted to them by a peripheral and flashing row of stars during a Posner cueing task. Similarly, studies that measured participants' brain responses via event-related potentials have shown that only voluntary attention resulted in brain activity representative of later and active stages of information processing into memory [35, 46, 47]. It has been suggested that voluntary attention improves memory of ROIs as it involves an active (top-down) effort in manipulating and storing information in cortical brain areas [14, 21, 70]. In contrast, involuntary attention provides a less robust form of memory encoding, as it involves stimulus driven (bottom-up) processing of information in parietal and temporal cortices and the brainstem [58], which appear to be less evolved [55] and more active during unsuccessful memory retrieval [70].

No CVR research has yet compared differences between central and peripheral cues in regard to memory. However, Rothe and colleagues [65] assessed participants' explicit memory after being guided to multiple static ROIs in a panoramic video with a subtle and overt peripheral flicker. If ROIs were outside of the FOV, flickers occurred at one of the four edges of the HMD's display. Participants could only recall 37% of information in the overt and below 18% of information in the subtle condition. Their study shows that peripheral cues encouraged few explicit memories, supporting previous research and assumptions. Our study extends their work by comparing memory of ROIs between peripheral and central cues.

2.5 Engagement with Cues in CVR

Engagement is characterized by users' perceived quality of their experience with an application [54]. It plays a crucial role for attention and memory, as it encourages "cognitive, temporal, affective and behavioral investment" in an activity, and it consists of six factors: focused attention; perceived usability; aesthetic appeal; endurability; novelty; felt involvement [54]. However, no study has yet investigated engagement with peripheral and central cues in CVR. Arrows may be perceived as more engaging than flickers, as they involve voluntary responses and may elicit curiosity due to their symbolic nature while focusing attention and encouraging vigilance [34, 55].

2.6 Limitations of Previous Research

The above assumptions have primarily been explored in non-immersive and controlled experimental environments, using desktop computers. Although a lot of research with virtual and augmented reality has been conducted to evaluate various attention-guiding techniques (see section 2.2 and [66] for a detailed overview), none of the reviewed studies has investigated if traditional theories on attention and memory are applicable to CVR, particularly in regard to guiding attention to out-of-view ROIs. Many of the studies were also limited in their ecological validity, featuring artificial tasks (e.g. [27,62]). Moreover, existing CVR-research has primarily been conducted either with immersive panoramic images [25] or videos with static ROIs and no scene-changes [65], which make it relatively easy to guide attention [43]. Therefore, it is unclear how effective the cues are in more demanding/disorienting environments. By addressing these limitations, the present study explored the question *"To what extent do a central arrow and peripheral flicker differ in their effectiveness for guiding and focusing attention in CVR?"*.

2.7 Hypotheses

The paper's hypotheses are: **H1)** participants will reach ROIs more slowly with the arrow than the flicker, as the arrow engages attention voluntarily whereas flicker responses are reflexive; **H2)** fewer targets will be reached when following the arrow compared to the flicker, as undesired arrows can be ignored due to involving voluntary attention whereas responses to flickers are reflexive; **H3)** the arrow will encourage longer fixations on cued ROIs than the flicker; **H4)** the arrow will encourage better memory of ROIs than the flicker; and **H5)** the arrow will be perceived as more engaging than the flicker.

3 METHOD

3.1 Participants

This study adopted a within-subjects experimental design to ensure that emerging differences between cues were not confounded by individual differences (e.g. familiarity with virtual environments, exploration behaviors or motor skills). Convenience sampling was used to recruit 25 participants (10 = male, 15 = female) between the ages 19 to 32 (M = 24.40, SD = 3.42) via an advertisement in University College London's online student newsletter, myUCL. Individuals were allowed to participate if they had normal or corrected-to-normal vision and were able to see, hear and walk unaided for the study's duration. They were not allowed to participate if they suffered from photosensitive epilepsy or regular/strong migraines. Participants with glasses were asked to wear contact lenses to ensure that the HMD could calibrate their gaze. All individuals were compensated in cash for their travel expenses. As demonstrated in supplemental Fig. S1 and Fig. S2, participants' familiarity and interest in VR as well as the presented video content was approximately normally distributed. Unfortunately, the eye-tracking and memory data from one male participant could not be analyzed, as the HMD collected incomplete gaze data due to a system error.

An initial pilot study was conducted with nine different participants to assess the appropriateness and clarity of the experimental materials, measures and procedure.

3.2 Materials

3.2.1 Apparatus

The FOVE-DK-0 HMD was used due to its infrared eye-tracking sensors, which have a frame rate of 120fps. The device uses an OLED display with a diagonal FOV of 100-degrees and has a resolution of 2560x1440 and a 70Hz refresh rate [19]. We did not correct for inertial measurement unit (IMU) drift as the "forward" direction reset at the start of each video and the accumulated drift was negligible. A pair of BOSE headphones played the videos' sound to participants on a medium volume. The cables of the HMD and

Figure 1: Experimental setup.

headphones were extended from the ceiling to ensure that participants could freely rotate in their revolving chair (see Fig. 1). Using Unity 5.6.1f1 (64 bit), software was created that allowed the cues to appear in real-time in the videos based on the user's gaze direction and HMD position. The software ran on a Windows 10 Pro desktop computer (64 bit) with an Intel i7-6700 CPU, which ran at 3.40 GHz with 32 GB of RAM. The video card was a NVIDIA GeForce GTX 1080.

3.2.2 Video Stimuli

Three panoramic videos from *YouTube* with 4K resolution were selected. The first video of a Celtic village [64] served as a two-minute introduction. It featured concurrent multiple ROIs, which remained static and, thus, allowed participants to get acquainted with the medium and cues without distraction. The videos "School of Rock" (5 min 10 s) [53] and "Pokémon" (4 min 16 s) [61] were used as the main experimental stimuli and were presented in a fixed order. They were adapted from Lin and colleagues' study [43], as they were validated in a similar research context and as they are representative of common multi-ROI panoramic experiences. "Pokémon" (henceforth referred to as ***exploration video***) is an example of exploratory panoramic videos, as it features various static ROIs and includes seven scene-changes. "School of Rock" (henceforth referred to as ***story video***) is representative of CVR experiences aimed at entertaining users via a central storyline with dynamic and static ROIs. The two experimental videos were chosen to increase the generalizability of collected eye-tracking data. They were not used for comparison, as order/learning effects may have confounded differences between them due to the fixed video order.

3.2.3 Regions of Interest (ROIs)

The practice video contained five ROIs. The exploration video and story video each contained seven ROIs to which participants were guided for eleven seconds each. ROIs were selected based on six principles: each ROI performs an at least 8-second action; each cue appears 3 seconds before the ROI's action; attention is directed to each ROI only once; the ROI does not move during the cue; no other video features guide attention to ROIs; directly consecutive ROIs appear on opposite sides of the video sphere. These principles were partly derived from the pilot study. All ROIs would have appeared outside of the user's FOV if he/she followed the cues without additional exploration. However, as users were free to observe the scene, some ROIs may have appeared within their FOV. While the cues in the practice video guided participants to ROIs consecutively, the ROIs in the exploration video and story video were distributed at varying intervals to prevent conditioned responses. Each ROI's rectangular size and position in pixels were defined for every 100 milliseconds of its eleven second-long duration to ensure that the cues guided users' attention to the correct location.

3.2.4 Cue Stimuli

Participants were presented with two cues (i.e. flicker/arrow) to guide their attention to ROIs. Both cues were made with Photoshop Elements 11. The study adapted a flicker, which is frequently used in the field (see Fig. 2). It was originally developed by Bailey and colleagues' [2] for guiding attention in images and further extended to virtual environments [26, 65]. Two black and white filled circles alternated at a frequency of 10Hz with a color intensity of [0,1]. The flicker's circular area covered 3.7% of the user's FOV in the HMD. Therefore, it was larger compared to previous research to make it clearly noticeable in the user's periphery. This was important, as previous research [26, 65] indicated that subtle peripheral cues are often overlooked, which would have placed our central cue at an advantage, causing bias. The outer 60% of the flicker were gradually smoothed, using up to 50% opacity. The flicker appeared three seconds before the ROI performed its eight-second action. It appeared at the ROI's center if the region was within the user's FOV. If the ROI was outside of the user's FOV, the flicker appeared 18% away from the HMD's edge that was closest to the ROI until the user rotated their head enough that the ROI fell within the FOV of the display. The flicker's dimensions and location were based on findings from the pilot study. The flicker disappeared if the user's gaze was within 18-degrees visual angle of the respective ROI's center. This measure of visual angle was chosen, as it signifies the edge of the eye's macula, which acts as the bridge between peripheral and central vision [36, 50].

The arrow's diameter and colors were similar to the flicker to make both cues as comparable as possible and to control for confounding effects. Although foveal vision is characterized by color perception whereas peripheral vision is predominantly black and white, a difference in cue color could have influenced users' engagement, attention to the video and willingness to observe ROIs. Due to the arrow's central position, it was considered sufficiently noticeable, which was supported by findings from our pilot study. The following approximate indications of the arrow's dimensions are based on scenarios where it pointed either to the left or right. Its horizontal diameter covered 10.4% of the x-axis of the user's FOV in the HMD. Its height covered 8% of the y-axis. The thickness of the arrow's line was 1.7% on the y-axis of the user's FOV. 33% of the arrow's surface area was white whereas 67% of its area consisted of a black boarder. The respective color intensities were [0,1]. The outer 40% of the arrow's black boarder were gradually smoothened, using up to 50% opacity. Based on the eye tracking in the FOVE HMD, the arrow appeared at the midpoint between both eyes of the user to ensure robustness against potential eye tracking issues (e.g. both eyes may not converge due to slight inaccuracies of the eye tracker). The arrow disappeared when the user's gaze reached the respective ROI's center. Until then, it gradually followed his/her gaze via an interpolation between the current eye and arrow position based on the below equation. The speed of the interpolation was specified via trial and error to encourage a movement of the arrow that was relatively fast yet smooth:

$$\text{Arrow velocity}\left(\frac{deg}{s}\right) = 2 \cdot (\text{Angle between eyes and arrow}) \quad (1)$$

The arrow stopped moving when it was within 2-degrees of the eye's midpoint (i.e. when entering foveal vision). It started following the user's gaze again when it was 8-degrees away from the central gaze point, as the paracentral visual field, after which visual acuity drastically declines, ends at approximately that point [13]. Both thresholds smoothed the arrow's movement against ballistic saccades and ensured that it stayed within the central visual field.

Based on findings from the pilot study, both cues lead the user back to each ROI once when his/her gaze moved more than 18-degrees visual angle away from the ROI's center. The re-direction ensured that users could clearly identify ROIs even if they missed

Figure 2: The flicker at two time points in the exploration video.
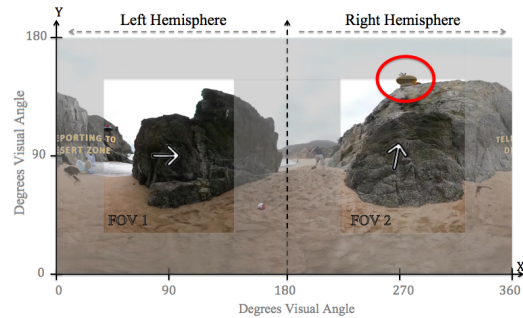

Figure 3: Arrow direction based on FOV and the ROI's (encircled in red) hemisphere.

them the first time after following the arrow or flicker. If an ROI appeared behind the user, the shortest path to it may have been over the top/bottom of the video sphere. However, to avoid leading participants over an unnatural path that may create discomfort, both cues first guided users to the correct hemisphere of the video's equirectangular projection before directing them along the shortest path to the ROI (see Fig. 3). The supplementary video to this paper displays both cues in action.

3.3 Measures and Scoring

This section outlines the study's dependent variables. The independent variable was the cue type: arrow versus flicker. All questionnaires were made with Qualtrics XM (see supplemental materials for the full question set and interview schedule).

3.3.1 Eye Tracking

The FOVE HMD recorded each participant's gaze position (i.e. x and y-coordinates) for every 100 milliseconds in the panoramic videos. If the participant gazed at the ROI with at least one eye, he/she was considered to have looked at it. The visual angle between the midpoint of the participant's left and right eye gaze at the time of a cue's onset and the respective ROI's center point was also calculated. Consequently, the average time per visual angle (s/deg) that each participant needed to reach ROIs after arrow and flicker onsets was calculated across the two experimental videos. The metric was adapted from Ben-Joseph and Greenstein [3], as it accommodates the impact of distance between user and ROI on the time needed to reach the latter. Furthermore, the average amount of time that participants looked at ROIs when guided by the arrow compared to the flicker was calculated. If a participant did not follow a cue or already looked at the ROI when it was supposed to appear, the respective data was excluded from analysis to avoid confounds.

3.3.2 Number of Followed Cues

To determine the cues' effectiveness in leading users to ROIs, the number of followed arrows and flickers was counted per participant across both videos. A cue was considered as "followed", if the participant's gaze was within the respective ROI at least once throughout the cue's duration.

3.3.3 Memory

Participants were presented with a recognition questionnaire, which assessed their memory of ROIs in the exploration video. It was refined based on results from the pilot study and contained three questions per ROI: "what did this Pokémon do?" (character-specific question); "who did x?" (behavior-specific question); "where was this Pokémon?" (context-specific question). Each question contained seven text and/or image answer options from which the correct answer(s) had to be selected. Using Adobe Photoshop CC (15.0), images of each Pokémon were cut from screenshots of the video, which were taken with the GoPro VR Player 3.0. Images of the Pokémon's original scene for context-specific questions were taken with the GoPro VR Player 3.0. Character-specific questions featured between one and three correct answers and behavior-specific questions contained between one and five correct answers. Context-specific questions only had one correct answer each. The questionnaire ended by assessing participants' knowledge of the name, appearance and qualities of each Pokémon before the study.

Each participant's response to character-specific and behavior-specific questions was given an accuracy score. Participants received one point for *selecting* the correct answer option and one point for each incorrect answer option, which they did *not select*. Therefore, the total possible score per question was 7. As only one answer could be selected for context-specific questions, participants received a score of one if they answered it correctly and received a score of zero if they answered incorrectly. Two mean accuracy scores per participant were then calculated for *each* question category, representing the participant's overall recognition ability when guided by the arrow and flicker respectively.

3.3.4 User Engagement

Participants' engagement with each cue was measured with the widely applied User Engagement Survey Short Form (UES-SF), which comprises four sub-scales: attentional focus (AF), perceived usability (PU), aesthetic appeal (AE), and reward value (RW). It was adopted due to its sub-scales' high reliability estimate, its short length and preserved content validity when questions are changed [54]. Three items from the long-form were added, based on pilot study insights: PU.3 ("I felt annoyed while using the cue"), PU7 ("I felt in control while using the cue"), and RW.5 ("I would recommend using the cue for similar VR videos to my family and friends").

The two questionnaires (one per cue type) were scored as per [54]: an average was calculated for each subscale by summing each participant's scores for relevant questions and dividing the sum by the number of items. Items PU-S.1, PU-S.2, PU.3, and PU-S.3 were reverse scored due to negative wording and overall engagement was calculated by averaging the sub-scales' scores.

3.3.5 Semi-Structured Interview

A ten to fifteen-minute semi-structured interview was conducted with each participant to collect deeper insights into their experiences with both cues. The schedule was structured into four parts, which inquired about: the participant's feelings about the cues, the cues' impact on attention and memory, the participant's cue preference, and observations during the main experimental procedure (e.g. "I noticed that you frequently did not follow the cue. Was there a reason for this?"). All interviews were audio-recorded and transcribed.

3.4 Experimental Procedure

The study took place in a Virtual Reality laboratory at University College London (UCL). Participants signed a consent form after reading an information sheet and were assured that they could withdraw at any time without mentioning a reason. They participated individually and sat next to the experimenter at a desk with a desktop computer and keyboard. Bottles of water were prepared in case participants experienced discomfort.

Participants completed an online questionnaire about their experience, interest, and familiarity with VR, Pokémon, and the film/musical School of Rock. They were then introduced to the FOVE HMD and were informed of the occasional presence of an arrow, which would be attached to their gaze, and a stimulus-based flicker in their periphery. Individuals were told that both cues might lead them toward information, which may or may not interest them. No task was assigned and participants could explore the videos of their own accord to ensure ecological validity. Their gaze was then calibrated and they were handed a pair of headphones. Thereafter, the three panoramic videos were played in a fixed order, starting with the practice video, followed by the story and the exploration video. The practice video contained two arrows and three flickers or vice versa, whereas the two experimental videos contained four arrows and three flickers or vice versa. The number of arrows and flickers in each video was counterbalanced across participants. Furthermore, the cue order in each video was randomized to control for order and learning effects.

After watching each video, participants were asked to stay within the HMD and to answer a few questions on signs of simulator sickness to ensure they were able to continue with the experiment. No participant reported any such signs. After all three videos, participants removed the HMD. On the desktop computer, they filled out the two UES-SFs. The order of the two questionnaires was counterbalanced across participants to control for order effects. Thereafter, participants completed the online recognition questionnaire. All of its questions occurred in a random order to control for effects of order and memory decay on recognition ability. Finally, the semi-structured interview was conducted and participants were debriefed. The study received ethical approval from the UCL Research Ethics Committee (Project ID: 4547/012).

4 RESULTS

4.1 Quantitative Data Treatment

The quantitative data was analyzed using IBM SPSS Statistics 25 and effect sizes were calculated and reported as per Cohen's [12] guidelines. As the time per degree that participants needed to reach ROIs as well as the number of followed cues violated the t-test assumption of normal distributions, two Wilcoxon Signed-Rank tests were performed on the data. The data for ROI observation time was analyzed with a paired-samples t-test, as all assumptions were met. Engagement and memory data consisted of average scores and could, thus, be treated as continuous [48]. They were analyzed via paired-samples t-tests.

4.2 Time per Degree to Reach ROIs

A Wilcoxon Signed-Rank Test was conducted to explore the effect of cue type on the time per degree that participants needed to reach ROIs. It displayed no statistically significant differences between the arrow (Md = .05) and flicker (Md = .03), z = -1.89, p = .06, with a medium effect size (r = .27). See Fig. S3 for a boxplot.

A close inspection of the raw data and interviews suggested that the flicker might have resulted in faster initial responses. A graph, showing the mean eye movement toward ROIs after cue-onset, is displayed in Fig. 4. It clearly demonstrates that the flicker created a stronger initial response, which was not sustained. Therefore, the average angular distance in the direction of ROIs that participants' gaze travelled within the first 500 ms after cue-onset was compared between the arrow and flicker post-hoc. The metric was inspired by previous research, which suggested that reflexive movements should be discernable from voluntary ones within the first 500 ms [2, 38, 52]. As the data met its assumptions, a paired-samples t-test was conducted and showed that the flicker encouraged users to travel 4.97-degrees visual angle farther, on average, within the specified time frame (M = 10.99, SD = 5.06) compared to the arrow (M = 6.02, SD = 3.29), t(23) = -3.65, p = .001. The eta squared statistic (η^2
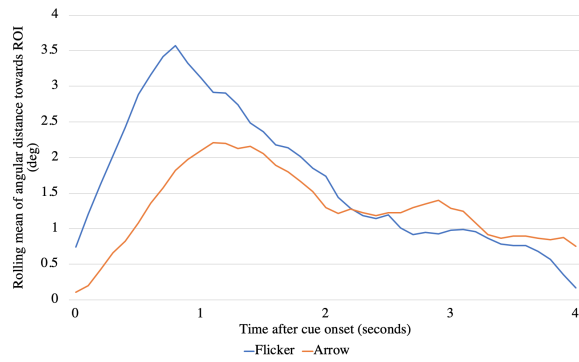
Figure 4: Mean angular distance moved by mid-point between eyes towards ROI following cue onset. To smooth the eye tracking data, a one second rolling average is shown (zero padded before cue onset).

Table 1: Results from the three memory questions.

Memory Question	Arrow M(SD)	Flicker M(SD)	Mean Difference	Lower CI	Upper CI
Character -specific	5.14 (.55)	5.17 (.70)	.03	-.33	.26
Behavior -specific	5.26 (.74)	5.22 (.75)	.05	-.20	2.96
Context -specific	.50 (.25)	.55 (.28)	.04	-.18	.10

= .37) revealed a large effect size and the 95% confidence interval ranged from -7.79 to -2.16. See Fig. S4 for a boxplot.

4.3 Number of Followed Cues

81.25% of the cues were followed and a Wilcoxon Signed-Rank Test revealed no statistically significant differences between the number of followed arrows (Md = 7.00) and flickers (Md =7.00), z = -.37, p = .71, with a small effect size (r = .05).

4.4 ROI Observation Time

A paired-samples t-test was performed to compare the impact of cue type on participants' ROI observation time. No statistically significant differences emerged between the arrow (M = 3.73, SD = .76) and the flicker (M = 3.83, SD = .95), t(23) = .52, p = .61, Mean diff. = .09, 95% CI from -.28 to .47, η^2 = .01 (small effect size).

4.5 Memory

Three paired-samples t-tests explored participants' recognition of ROI-related information when guided by the arrow compared to the flicker. They revealed no statistically significant differences between both cues for character-specific questions, t(23) = -.22, p = .83 (two-tailed), η^2 = .00 (small effect size). No differences between the cues also existed for behavior-specific questions, t(23) = .38, p = .71 (two-tailed), η^2 = .01 (small effect size). Finally, no differences emerged for context-specific questions, t(23) = -.61, p = .55 (two-tailed), η^2 = .00 (small effect size). See Table 1 for further information.

4.6 User Engagement

A paired-samples t-test was conducted to compare participants' over-all engagement with the arrow and the flicker. It showed no statistically significant differences between cues, t(24) = 1.62, p = .12 (two-tailed). However, the eta squared statistic revealed a moderate effect size (η^2 = .10). Four further planned exploratory paired-samples t-tests compared participants' scores on the four sub-scales of the UES-SF between the arrow and the flicker. The analyses

Table 2: Results on User Engagement on the UES-SF.

Engagement Level	Arrow M(SD)	Flicker M(SD)	Mean Difference	Lower CI	Upper CI
Overall engagement	3.51 (.64)	3.29 (.68)	.23	-.06	.52
Attentional focus	2.97 (.62)	2.93 (.68)	.04	-.27	.35
Perceived usability	3.85 (.67)	3.58 (.75)	.26	-.07	.60
Aesthetic appeal	3.13 (1.03)	2.97 (.97)	.16	-.38	.69
Reward value	3.79 (.76)	3.41 (.92)	.38	.02	.74

demonstrated no significant differences between cues on the attentional focus scale, t(24) = .27, p = .79 (two-tailed), η^2 = .00 (small effect size). No differences also emerged for the perceived usability scale, t(24) = 1.64, p =.11 (two-tailed), η^2 = .10 (moderate effect size). Furthermore, the cues did not differ on the aesthetic appeal scale, t(24) = .61, p = .55 (two-tailed), η^2 = .02 (small effect size). However, the arrow received significantly higher ratings than the flicker on the reward scale, t(24) = 2.19, p = .04 (two-tailed). The eta-squared statistic (η^2 = .17) emphasized this finding by revealing quite a large effect size. See Table 2 for further information.

4.7 Semi-Structured Interviews

The study's principal investigator first analyzed five randomly selected transcripts on their semantic content from which 41 codes were generated from the bottom up, using inductive thematic analysis as outlined by [4]. Codes were categorized into six preliminary themes and applied holistically across the remaining transcripts. They were updated iteratively until the final code-book contained 27 codes and five themes. Thereafter, another researcher, who was unfamiliar with the study and participants, independently reviewed the themes/codes in respect to six randomly selected transcripts. He agreed with all themes and 83.9% of codes. Disagreements were resolved through discussion and the final code-book contained 24 codes (see table S1). The below sub-sections outline the emerged themes, which are supported with interview quotes (see section 5.5 for an overview of key findings). 68% of participants preferred the arrow whereas the remaining 32% preferred the flicker.

4.7.1 Attention Grabbing Differences Between Cues

In line with quantitative findings, the majority of participants reported that "*with the arrow, [they] very much felt like [they] had to think about what [they were] doing first whereas the flicker was very much instinctive*" (P25), as its location was easily discernible and as it was often interpreted as a "*danger signal*" (P2). The flicker may have also grabbed attention faster initially as many participants were sensitive to flashing. Some even experienced adverse physical reactions, such as "*feeling almost a bit nauseous*" (P18), which is known as photophobia. Participant 24 compared "*it to when the sun is shining and it reflects on something (. . .) how sometimes it hurts your eyes*". However, substantial variation in participants' responses existed, as some individuals struggled to notice the flicker at all. P18 reported that she "*always noticed the flicker after two seconds and sometimes it just disappeared when [she] noticed it*". Many of these participants also reported that the flicker "*sometimes got a little bit lost in the video and the messiness*" (P22). They mentioned that negative effects of visual crowding on noticing flickers were especially pronounced at larger eccentricities from the center of vision on the horizontal axis. On one hand, this finding may suggest that some participants had weak left-and right-sided vision. On the other hand, participants could have experienced cognitive workload due

to being overwhelmed by the CVR setting, leading to more focused vision. Furthermore, some participants felt that the flicker's initial superiority disappeared with practice as they got used to the arrow. For instance, P20 mentioned that he did not get "*used to the arrow as quickly, because it was moving around and with [him] (...) and arrows usually don't do that*".

4.7.2 Guidance Features of Cues

Individuals who did not demonstrate adverse physical reactions to flickering reported "*the flicker is something that you can sort of zone out of and not pay attention to*" (P25) whereas "*it was almost as if, because it is an arrow, it's more of an important thing*" (P26). The arrow also guided attention more forcefully yet clearly due to its pointy shape and "*because it was literally in the [FOV's] center*" (P23). In contrast, the flicker often appeared subtle, vague and confusing, as it disappeared before users could focus on a ROI. Indeed, P9 described how she felt "*like the arrow makes you go with it smoothly. The flicker is more like (...) just a quick reaction, so [she] felt a little bit disoriented*". Nevertheless, individuals who preferred the flicker did not appreciate the arrow's forceful guidance, particularly its re-direction to ROIs, as "*it seems like the arrow came back more often (...) and so (...) [participants] felt a little bit punished at times for looking away too quickly*" (P20).

4.7.3 Exploration and Focus

Participants agreed that both cues "*directed your attention to something that you probably would have missed otherwise*" (P17) while distracting users from the video occasionally. However, in contrast to the flicker, the arrow encouraged participants to "*look around at the whole [360-degree] scene*" (P10) more, as it "*appeared to have a wider range of directions*" (P4) and as participants more "*slowly moved along with it*" (P11). Thus, it incited "*the curiosity to explore more*" (P16). Furthermore, most participants reported that the arrow made them focus more on exact ROIs due to its pointed shape whereas the flicker encouraged greater exploration of ROIs' *immediate* surroundings, as it was circular and encouraged more rapid/disorienting head-turns. For example, P4 reported that with the arrow "*it's just that tip you see, so you know exactly what you're meant to be looking at and focusing on, whereas the flicker by nature has a circular shape, so you don't know where within the area to focus*" (P4). As this finding was not reflected in the quantitative analysis, the arrow may have caused participants to overestimate their focus on ROIs, making it appear as more rewarding.

4.7.4 Importance of Context for Cue Success

Participants agreed that video genre influenced the effectiveness and appropriateness of both cues. They agreed that the arrow was particularly useful for search-based tasks, exploratory or educational CVR-experiences due to its forceful and clear guidance, encouraging ROI-focus and scene exploration. In contrast, the flicker seemed more appropriate for entertaining or immersive experiences in which cues should not distract from the storyline. P26 explained "*if it was like, an educational video, I would use an arrow, because (...) the attention would be more on the object, whereas if I was trying to immerse someone in something (...) like have them watch VR-TV or a game or something like that, then I would use a flash, because it's a lot less: Here is something you NEED to look at*". However, participants also agreed that neither of the cues is needed in panoramic videos, which do not include scene changes/teleportation, as they can negatively impact video enjoyment. For example, P8 "*was confused what to look for [in the introduction video], because [she had] already looked around*" the scene without cue guidance while P24 "*felt slightly annoyed*" with the cues in this context.

Moreover, many participants emphasized the importance of making cues optional, as their effectiveness often depended on features of the user's personality and expectation. Indecisive individuals who were "*very used to follow instructions*" (P5) liked the cues and "*whenever a cue popped up, [they] would look at that [ROI]*" (P5). In contrast, individuals who disliked instructions felt that "*it would be more rewarding to find something by accident rather than being guided to it*" (P18). They ignored cues or immediately returned to their previous position in the scene, as "*there were other things in the video around [them] that were more interesting*" (P8).

4.7.5 Memory of ROIs

In line with this study's quantitative findings, the majority of participants suggested that "*both (cues) definitely helped [them] remember but only when [they were] interested in what [they were] pointed to*" (P5). Eight participants believed that "*most of [their] memories were from the arrow, because the arrow can lead you to a point and your first impression is that point*" (P7) and "*because it was out of context*" (P10). Thus, the arrow encouraged some users to be more certain of their ROI-related memory compared to the flicker. Additionally, some participants implied that the "*arrow helped [them] understand and remember the whole story*" (P15) as they explored more with it, making the experience appear more worthwhile.

5 Discussion

This study investigated the effectiveness of a central cue (i.e. arrow) compared to a peripheral cue (i.e. flicker) for guiding and focusing attention in multi-ROI panoramic videos. It was the first in its field to compare both cues in regard to users' reaction and observation times, ROI-related memory, engagement and overall user experience.

5.1 Time per Degree to Reach ROIs

Our first hypothesis was not entirely confirmed, as no statistically significant differences emerged between the cues for the total time that participants needed to reach ROIs. However, the flicker resulted in significantly faster initial movements toward ROIs within the first 500 ms. Although, this finding suggests that the flicker encouraged a more automatic attentional shift than the arrow, it did not appear to induce a reflexive saccadic response, as participants covered, on average, only 11 degrees visual angle within the time frame [5].

One explanation for the above findings could be that following flickers to out-of-view ROIs required voluntary attention even though initial reactions were involuntary. Therefore, flickers could be more effective than arrows for guiding attention to in-view ROIs. Indeed, Brodal [6] suggested that voluntary attention and activation of higher motor centers in the brain are required for postural control, locomotion and adjusting gaze precision. Moreover, our participants reported that they could ignore the flicker even though their initial reaction was to look in its direction. As most ROIs appeared outside of participants' FOV, no statistically significant difference between cues may have appeared overall.

Alternatively, the flicker may have encouraged reflexive saccades only for some participants. Indeed, many individuals have weak peripheral vision and objects become even less distinguishable at greater eccentricities from the center and with increased visual crowding [24], which has also been reported in the interviews. Another explanation could be that some participants may have been overwhelmed by the immersive CVR experience and, thus, focused their vision more on central objects in their FOV. Therefore, many participants may have missed flickers or responded with delays to which saccadic suppression (i.e. temporary blindness during eye movement) may have further contributed.

A third explanation could be that participants became desensitized to the flicker with time [20, 33], which also explains why many reported an ability to ignore it. Desensitization refers to decreased sensitivity of eye receptor cells due to constant stimulation, which can turn reflexive responses voluntary [41]. Subconsciously perceived flickering of the HMD display may have further increased the likelihood of desensitization.

Finally, as this study used a noticeable peripheral cue in contrast to subtler versions in previous research, for example [25, 26], it is possible that the cue's boldness encouraged more voluntary reactions, reducing differences between cues in our analyses. Further research should explore this prospect.

5.2 Followed Cues, Observation Time and Memory

The study's second hypothesis was not met, as no significant differences existed between the number of followed arrows and flickers. As suggested, it can be assumed that participants actively decided to follow both cues equally due to voluntary attention components.

The arrow was also predicted to encourage longer observations of ROIs due to inciting greater interest based on its elicited voluntary attention compared to the flicker [35, 55]. Moreover, the arrow was expected to increase memory due to engaging top-down processes in contrast to the flicker, which would engage bottom-up processing [31, 35]. However, no significant differences emerged between cues for both measures. Indeed, most participants reported that they remained at cued locations only if ROIs engaged their interest. The finding appears to confirm that both cues may have engaged voluntary attention.

5.3 User Engagement

Contrasting the fourth hypothesis, there were no statistically significant differences between cues regarding participants' overall engagement. However, the arrow was rated as more rewarding. Unfortunately, the UES-SF only asked a small number of questions, which did not assess participants on the most crucial factors that distinguished both cues according to the interviews (i.e. cue precision, accuracy, visibility and effects on well-being). Further research could explore these features more.

5.4 Methodological Limitations and Further Work

This study has a number of limitations. First, as it only explored participants' responses to two multi-ROI videos and fourteen cues in total, further larger-scale research is needed to assess the generalizability of its findings to a range of different video types, genres, and ROIs [10]. In a similar vein, the findings' generalizability to HMDs with a larger FOV than FOVE-DK-0 could be explored, as such HMDs might need fewer peripheral flickers to guide attention to out-of-view ROIs (i.e. faster attention guidance). Second, a sampling bias existed, as only university students participated, who tend to have medium to high socioeconomic status [69] and represent a narrow age group at the peak of its cognitive abilities [28]. Third, the study had slightly limited ecological validity, as the experimental procedure and laboratory environment may have encouraged participants to pay more attention to the cues and videos compared to at-home CVR-consumption [10]. Finally, the use of a fixed video order prevented the researchers from investigating differences between the two experimental videos, as the internal validity of such analyses would have been confounded by order and learning effects. However, this was necessary to standardize the procedure. Indeed, if questionnaires had been presented directly after each video while counterbalancing the video order, participants could have been aware of the study's goal. They may have been encouraged to pay more attention to cues and ROIs in the following video, which would not accurately reflect how they would normally engage with the experience (i.e. response bias). Equally, all questionnaires were not presented after both videos were shown in a counterbalanced order, as this approach would have resulted in longer intervals between the exploration video and recognition questionnaire for half of the sample. Finally, further research should investigate cue-related differences between videos and the potential impact of learning effects. It should also explore the cues' impact on users' sense of presence, distraction and annoyance to give a fuller picture of the user experience.

5.5 Design Implications

The study implies that existing theories of attention and memory (outlined in sections 2.3-4) are not entirely applicable to CVR. Therefore, existing taxonomies for directing attention in CVR [66] may need to be adjusted to inform the design of future methods. The quantitative findings suggest that, on average, both cues are effective in guiding and focusing attention as well as encouraging users to memorize ROI-related information. However, users preferred the arrow, which appeared to be most suitable for educational or exploration-based videos due to its forceful appearance. Some participants deemed the flicker more suitable for entertaining experiences due to its playful nature although responses were highly dependent on individual differences. While some participants found it difficult to spot flickers, others experienced extreme eye sensitivities (including photophobia), which has not been reported in previous research. Therefore, although filmmakers could advise users on a particular cue based on video genre/type, consumers should be warned about flicker side effects and be allowed to choose/change cues. Participants also reported that they needed to learn how to use the arrow first due to its unusual gaze attachment, which supports findings from previous CVR-research [62]. Filmmakers thus need to adjust for learning effects when using similar arrows, for example by integrating playful introduction videos. Finally, videos containing single ROIs, no movement as well as no scene changes may not require cues and qualitative findings indicated that cues should be used sparsely in general to increase the user's agency and enjoyment.

6 Conclusion

This study investigated the effectiveness of using central versus peripheral cues in multi-ROI panoramic videos. Participants' gaze travelled a greater distance toward ROIs within the first 500 ms after flicker- compared to arrow-onsets. However, no significant differences between cues emerged in regard to the total time taken to reach and observe ROIs, memory and user engagement. Nevertheless, participants rated the arrow as more rewarding and revealed a preference for it. The findings contradict previous theoretical assumptions and existing taxonomies. Indeed, our results imply that peripheral cues may engage voluntary attention components due to the large distances involved in panoramic videos. The study suggests that overall cue success depends on the user's intentions and it emphasizes the importance of making cues optional. Furthermore, it implies that different cues may be more appropriate in different video contexts. However, due to the narrow age group under investigation and the relatively small number of videos and cues/ROIs in this study, its results should be interpreted with caution. Further research is needed to explore different user populations with varying VR-expertise and strength of vision and to assess the findings' generalizability to a broader range of panoramic videos and genres. Finally, as this study's flicker does not appear universally suitable due to occasional photophobic side effects, its design may need to be altered.

References

[1] J. Abrams, A. Barbot, and A. Carrasco. Voluntary attention increases perceived spatial frequency. *Attention, Perception, & Psychophysics*, 72(6):1510–1521, Aug. 2010. doi: 10.3758/APP.72.6.1510

[2] R. Bailey, A. McNamara, N. Sudarsanam, and C. Grimm. Subtle gaze direction. *ACM Transactions on Graphics*, 28(4):100:1–100:14, Sept. 2009. doi: 10.1145/1559755.1559757

[3] E. Ben-Joseph and E. Greenstein. Gaze direction in virtual reality using illumination modulation and sound. Stanford University, 2016.

[4] V. Braun and V. Clarke. *Successful qualitative research: A practical guide for beginners*. SAGE Publications, London, England, 2013.

[5] E. Britannica. Sensory reception, human vision, structure and function of the human eye. In *The new encyclopedia britannica*, vol. 27. Encyclopedia Britannica Inc., Chicaco, USA, 1987.

[6] P. Brodal. *The central nervous system*. Oxford University Press, Oxford, England, 2010.

[7] A. Brown, A. Sheikh, M. Evans, and Z. Watson. Directing attention in 360-degree video. In *IBC 2016 Conference*. IBC. doi: 10.1049/ibc.2016.0029

[8] A. Brown, J. Turner, J. Patterson, A. Schmitz, M. Armstrong, and M. Glancy. Subtitles in 360-degree video. In *Adjunct Publication of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video*, TVX '17 Adjunct, pp. 3–8. ACM, 2017. doi: 10.1145/3084289.3089915

[9] C. Brown, G. Bhutra, M. Suhail, Q. Xu, and E. D. Ragan. Coordinating attention and cooperation in multi-user virtual reality narratives. VR, pp. 377–378, 2017. doi: 10.1109/VR.2017.7892334

[10] A. Bryman. *Social research methods*. Oxford University Press, Oxford, England, 2012.

[11] M. Carrasco and B. McElree. Covert attention accelerates the rate of visual information processing. *Proceedings of the National Academy of Sciences of the United States of America*, 98(9):5363–5367, April 2001. doi: 10.1073/pnas.081074098

[12] J. Cohen. *Statistical power analysis for the behavioral sciences*. Lawrence Earlbaum Associates, Hillsdale, NJ, USA, 2 nd ed., 1988.

[13] A. M. Colman. *A dictionary of psychology*. Oxford University Press, Oxford, England, 3 rd ed., 2008.

[14] M. Corbetta, J. M. Kincade, J. M. Ollinger, M. P. McAvoy, and G. L. Shulman. Voluntary orienting is dissociated from target detection in human posterior parietal cortex. *Nature Neuroscience*, 3(3):292–297, March 2000. doi: 10.1038/73009

[15] X. Corbillon, G. Simon, A. Devlic, and J. Chakareski. Viewport-adaptive navigable 360-degree video delivery. In *2017 IEEE International Conference on Communications (ICC)*, pp. 1–7. IEEE, May 2017. doi: 10.1109/ICC.2017.7996611

[16] F. Danieau, A. Guillo, and R. Doré. Attention guidance for immersive video content in head-mounted displays. In *2017 IEEE Virtual Reality (VR)*, pp. 205–206. IEEE, March 2017. doi: 10.1109/VR.2017.7892248

[17] M. Eimer. The time course of spatial orienting elicited by central and peripheral cues: Evidence from event-related brain potentials. *Biological Psychology*, 53(2):253–258, 2000. doi: 10.1016/S0301-0511(00)00049-1

[18] A. Elmezeny, N. Edenhofer, and J. Wimmer. Immersive storytelling in 360-degree videos: An analysis of interplay between narrative and technical immersion. *Journal of Virtual Worlds Research*, 11(1), April 2018.

[19] FOVE. FOVE 0 eye tracking VR devkit for developers, creators, researchers. https://www.getfove.com, Accessed on 13 Aug., 2019.

[20] M. Gazzaniga, R. B. Ivry, and G. R. Mangun. *Cognitive neuroscience: The biology of the mind (International Student Edition)*. W. W. Norton & Company, New York, NY, USA, 4 th ed., 2014.

[21] C. D. Gilbert and M. Sigman. Brain states: Top-down influences in sensory processing. *Neuron*, 54(5):677–696, June 2007. doi: 10.1016/j.neuron.2007.05.019

[22] A. M. Giordano, B. McElree, and M. Carrasco. On the automaticity and flexibility of covert attention: a speed-accuracy trade-off analysis. *Journal of vision*, 9(3):1–10, March 2009. doi: 10.1167/9.3.30

[23] F. Gobet, P. Chassy, and M. Bilalic. *Foundations of cognitive psychology*. McGraw-Hill Education, Berkshire, UK, Jan. 2011.

[24] J. A. Greenwood, M. Szinte, B. Sayim, and P. Cavanagh. Variations in crowding, saccadic precision, and spatial localization reveal the shared topology of spatial vision. *Proceedings of the National Academy of Sciences*, 114(17):E3573, April 2017. doi: 10.1073/pnas.1615504114

[25] S. Grogorick, G. Albuquerque, J.-P. Tauscher, and M. Magnor. Comparison of unobtrusive visual guidance methods in an immersive dome environment. *ACM Transactions on Applied Perception*, 15(4):27:1–27:11, Sept. 2018. doi: 10.1145/3238303

[26] S. Grogorick, M. Stengel, E. Eisemann, and M. Magnor. Subtle gaze guidance for immersive environments. In *Proceedings of the ACM Symposium on Applied Perception*, SAP '17, pp. 4:1–4:7. ACM, 2017. doi: 10.1145/3119881.3119890

[27] U. Gruenfeld, A. El All, S. Boll, and W. Heuten. Beyond halo and wedge: visualizing out-of-view objects on head-mounted virtual and augmented reality devices. In *Proceedings of the 20th International Conference on Human-Computer Interaction with Mobile Devices and Services*, MobileHCI '18, pp. 1–11. ACM, New York, NY, USA, 2018. doi: 10.1145/3229434.3229438

[28] B. Guerra-Carrillo, K. Katovich, and S. A. Bunge. Does higher education hone cognitive functioning and learning efficacy? findings from a large and diverse sample. *PLOS ONE*, 12(8):e0182276, Aug. 2017. doi: 10.1371/journal.pone.0182276

[29] E. Guervós, J. J. Ruiz, P. Pérez, J. A. Muñoz, C. Díaz, and N. García. Using 360 VR video to improve the learning experience in veterinary medicine university degree. *Electronic Imaging*, 2019(12):217–1–217–7, 2019.

[30] H. Hata, H. Koike, and Y. Sato. Visual guidance with unnoticed blur effect. In *Proceedings of the International Working Conference on Advanced Visual Interfaces*, AVI '16, pp. 28–35. ACM, June 2016. doi: 10.1145/2909132.2909254

[31] B. J. A. Hauer and C. M. MacLeod. Endogenous versus exogenous attentional cuing effects on memory. *Acta Psychologica*, 122(3):305 – 320, 2006. doi: 10.1016/j.actpsy.2005.12.008

[32] R. C. Herault, A. Lincke, M. Milrad, E.-S. Forsgärde, and C. Elmqvist. Using 360-degrees interactive videos in patient trauma treatment education: design, development and evaluation aspects. *Smart Learning Environments*, 5(1):26, Oct. 2018. doi: 10.1186/s40561-018-0074-x

[33] M. D. Hilchey, R. M. Klein, and J. Satel. Returning to "inhibition of return" by dissociating long-term oculomotor IOR from short-term sensory adaptation and other nonoculomotor "inhibitory" cueing effects. *Journal of experimental psychology: Human perception and performance*, 40(4):1603–16, 2014.

[34] E. M. Hitchcock, W. N. Dember, J. S. Warm, B. W. Moroney, and J. E. See. Effects of cueing and knowledge of results on workload and boredom in sustained attention. *Human Factors*, 41(3):365–372, Sept. 1999. doi: 10.1518/0018720099779610987

[35] J. Hopfinger and E. L Parks. Involuntary attention. In G. R. Mangun, ed., *The neuroscience of attention: Attentional control and selection*, pp. 30–53. Oxford University Press, New York, NY, USA, 2012.

[36] J. M. Johnston. *Eye studies: A series of lessons on vision and visual tests*. J. M. Johnston, Chicaco, USA, 1892.

[37] J. Jonides. Voluntary versus automatic control over the mind's eye's movement. In J. Long and A. Baddeley, eds., *Attention and performance IX*. Erlbaum, Hillsdale, NJ, USA, 1981.

[38] P. Kurtz, K. A. Shapcott, J. Kaiser, J. T. Schmiedt, and M. C. Schmid. The influence of endogenous and exogenous spatial attention on decision confidence. *Scientific reports*, 7(1):6431–6431, July 2017. doi: 10.1038/s41598-017-06715-w

[39] A. Kvisgaard, S. Klem, T. L. Nielsen, E. I. Rafferty, N. C. Nilsson, E. R. Høeg, and R. Nordahl. Frames to zones: Applying mise-en-scene techniques in cinematic virtual reality. In *2019 IEEE 5th Workshop on Everyday Virtual Reality*. IEEE, March 2019.

[40] L. K. Langley, C. K. Friesen, A. L. Saville, and A. T. Ciernia. Timing of reflexive visuospatial orienting in young, young-old, and old-old adults. *Attention, perception & psychophysics*, 73(5):1546–1561, July 2011. doi: 10.3758/s13414-011-0108-8

[41] A. Lim, V. Eng, S. M. J. Janssen, and J. Satel. Sensory adaptation and inhibition of return: dissociating multiple inhibitory cueing effects. *Experimental Brain Research*, 236(5):1369–1382, May 2018. doi: 10.1007/s00221-018-5225-3

[42] 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. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*, CHI '17, pp. 2535–2545. ACM, 2017. doi: 10.1145/3025453.3025757

[43] Y.-T. Lin, Y.-C. Liao, S.-Y. Teng, Y.-J. Chung, L. Chan, and B.-Y. Chen. Outside-in: Visualizing out-of-sight regions-of-interest in a 360 video using spatial picture-in-picture previews. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*,

UIST '17, pp. 255–265. ACM, 2017. doi: 10.1145/3126594.3126656

[44] X. Liu, Q. Xiao, V. Gopalakrishnan, B. Han, F. Qian, and M. Varvello. 360-degree innovations for panoramic video streaming. In *Proceedings of the 16th ACM Workshop on Hot Topics in Networks*, HotNets-XVI, pp. 50–56. ACM, 2017. doi: 10.1145/3152434.3152443

[45] A. MacQuarrie and A. Steed. Cinematic virtual reality: Evaluating the effect of display type on the viewing experience for panoramic video. In *2017 IEEE Virtual Reality (VR)*, pp. 45–54. IEEE, March 2017. doi: 10.1109/VR.2017.7892230

[46] G. R. Mangun and S. A. Hillyard. Modulations of sensory-evoked brain potentials indicate changes in perceptual processing during visual-spatial priming. *Journal of Experimental Psychology: Human Perception and Performance*, 17(4):1057–1074, 1991. doi: 10.1037/0096-1523.17.4.1057

[47] A. Martinez, F. DiRusso, L. Anllo-Vento, M. I. Sereno, R. B. Buxton, and S. A. Hillyard. Putting spatial attention on the map: timing and localization of stimulus selection processes in striate and extrastriate visual areas. *Vision Research*, 41(10):1437–1457, May 2001. doi: 10.1016/S0042-6989(00)00267-4

[48] J. H. McDonald. *Handbook of biological statistics*. Sparky House Publishing, Baltimore, MD, USA, 3 rd ed., 2014.

[49] M. Millodot. *Dictionary of optometry and visual science*. Elsevier Health Sciences, London, UK, 8 th ed., 2014.

[50] C. G. Morris. *Academic press dictionary of science and technology*. Academic Press, Cambridge, MA, USA, 1 ed., 1992.

[51] H. J. Müller and J. M. Findlay. The effect of visual attention on peripheral discrimination thresholds in single and multiple element displays. *Acta Psychologica*, 69(2):129–155, 1988. doi: 10.1016/0001-6918(88)90003-0

[52] H. J. Müller and P. M. Rabbitt. Reflexive and voluntary orienting of visual attention: Time course of activation and resistance to interruption. *Journal of Experimental Psychology: Human Perception and Performance*, 15(2):315–330, 1989. doi: 10.1037/0096-1523.15.2.315

[53] S. of Rock the Musical. *SCHOOL OF ROCK: The Musical - "You're in the Band" (360 Video)*. https://www.youtube.com/watch?v=GFRPXRhBYOI, Oct. 14, 2015 (accessed on Aug. 13, 2019).

[54] H. L. O'Brien, P. Cairns, and M. Hall. A practical approach to measuring user engagement with the refined user engagement scale (UES) and new UES short form. *International Journal of Human-Computer Studies*, 112:28–39, April 2018. doi: 10.1016/j.ijhcs.2018.01.004

[55] Y. Pinto, A. R. van der Leij, I. G. Sligte, V. A. F. Lamme, and H. S. Scholte. Bottom-up and top-down attention are independent. *Journal of Vision*, 13(3):16–16, 2013. doi: 10.1167/13.3.16

[56] M. Posner. Orienting of attention. *The Quarterly Journal of Experimental Psychology*, 32(1):3–25, 1980. doi: 10.1080/00335558008248231

[57] M. I. Posner and Y. Cohen. Components of visual orienting. In H. Bouma and D. G. Bouwhuis, eds., *Attention and performance*, vol. 32, pp. 531–556. Erlbaum, Hillsdale, NJ, 1984.

[58] M. I. Posner and S. E. Petersen. The attention system of the human brain. *Annual Review of Neuroscience*, 13(1):25–42, March 1990. doi: 10.1146/annurev.ne.13.030190.000325

[59] M. I. Posner, C. R. Snyder, and B. J. Davidson. Attention and the detection of signals. *Journal of Experimental Psychology: General*, 109(2):160–174, 1980. doi: 10.1037/0096-3445.109.2.160

[60] Y. Pulijala, M. Ma, and A. Ayoub. VR surgery: Interactive virtual reality application for training oral and maxillofacial surgeons using oculus rift and leap motion. In M. Ma and A. Oikonomou, eds., *Serious Games and Edutainment Applications*, vol. 2, pp. 187–202. Springer Internation Publishing, Cham, Switzerland, 2017. doi: 10.1007/978-3-319-51645-5_8

[61] F. Rad. *Pokèmon 360 - CATCH 'EM ALL IN VR!* https://www.youtube.com/watch?v=pHUVS_GrIeM, Sep. 13, 2016 (accessed on Aug. 13, 2019).

[62] P. Renner and T. Pfeiffer. Attention guiding techniques using peripheral vision and eye tracking for feedback in augmented-reality-based assistance systems. *2017 IEEE Symposium on 3D User Interfaces (3DUI)*, pp. 186–194, 2017.

[63] L. Riggio and K. Kirsner. The relationship between central cues and peripheral cues in covert visual orientation. *Perception & Psychophysics*, 59(6):885–899, Jan. 1997. doi: 10.3758/BF03205506

[64] S. Rothe. *Keltisches Dorf Stöffling*. https://www.youtube.com/watch?v=A9yF_DH5sEY, Nov. 16, 2016 (accessed on Aug. 13, 2019).

[65] S. Rothe, F. Althammer, and M. Khamis. GazeRecall: Using gaze direction to increase recall of details in cinematic virtual reality. In *Proceedings of the 17th International Conference on Mobile and Ubiquitous Multimedia*, MUM 2018, pp. 115–119. ACM, 2018. doi: 10.1145/3282894.3282903

[66] S. Rothe, D. Buschek, and H. Hußmann. Guidance in cinematic virtual reality-taxonomy, research status and challenges. *Multimodal Technologies and Interaction*, 3(1), 2019. doi: 10.3390/mti3010019

[67] S. Rothe, H. Hußmann, and M. Allary. Diegetic cues for guiding the viewer in cinematic virtual reality. In *Proceedings of the 23rd ACM Symposium on Virtual Reality Software and Technology*, VRST '17, pp. 54:1–54:2. ACM, 2017. doi: 10.1145/3139131.3143421

[68] Y. Sato, Y. Sugano, A. Sugimoto, Y. Kuno, and H. Koike. Sensing and controlling human gaze in daily living space for human-harmonized information environments. In T. Nishida, ed., *Human-Harmonized Information Technology, Volume 1: Vertical Impact*, pp. 199–237. Springer Japan, Tokyo, 2016. doi: 10.1007/978-4-431-55867-5_8

[69] S. Thomson. Achievement at school and socioeconomic background—an educational perspective. *Npj Science of Learning*, 3(1):5, March 2018. doi: 10.1038/s41539-018-0022-0

[70] M. R. Uncapher, J. B. Hutchinson, and A. D. Wagner. Dissociable effects of top-down and bottom-up attention during episodic encoding. *The Journal of neuroscience : the official journal of the Society for Neuroscience*, 31(35):12613–12628, Aug. 2011. doi: 10.1523/JNEUROSCI.0152-11.2011

[71] A. van Hoff. Virtual reality and the future of immersive entertainment. In *Proceedings of the 2017 ACM International Conference on Interactive Experiences for TV and Online Video*, TVX '17, pp. 129–129. ACM, 2017. doi: 10.1145/3077548.3080529