RadarVR: Exploring Spatiotemporal Visual Guidance in Cinematic VR

Sean J. Liu*
Reality Labs Research, Meta
Stanford University

Rorik Henrikson†
Reality Labs Research, Meta

Tovi Grossman‡
University of Toronto

Michael Glueck†
Reality Labs Research, Meta

Mark Parent†
Reality Labs Research, Meta

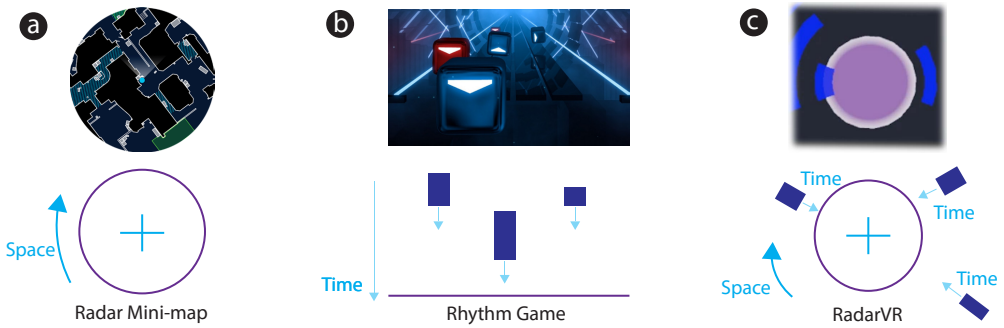

Figure 1: RadarVR provides spatiotemporal visual guidance in cinematic virtual reality (i.e., 360° videos). (a) Radar mini-maps are used for spatial navigation in games [58]. (b) Rhythm games encode temporal information as distance from a reference line [26]. (c) RadarVR blends these familiar design metaphors and visualizes regions of interest (ROIs) in space and time around a radar for visual guidance. RadarVR offers viewers a look-ahead time and allows them to plan their head motion in advance of upcoming ROIs.

ABSTRACT

In cinematic VR, viewers can only see a limited portion of the scene at any time. As a result, they may miss important events outside their field of view. While there are many techniques which offer spatial guidance (where to look), there has been little work on temporal guidance (when to look). Temporal guidance offers viewers a look-ahead time and allows viewers to plan their head motion for important events. This paper introduces spatiotemporal visual guidance and presents a new widget, *RadarVR*, which shows both spatial and temporal information of regions of interest (ROIs) in a video. Using RadarVR, we conducted a study to investigate the impact of temporal guidance and explore trade-offs between spatiotemporal and spatial-only visual guidance. Results show spatiotemporal feedback allows users to see a greater percentage of ROIs, with 81% more seen from their initial onset. We discuss design implications for future work in this space.

CCS CONCEPTS

• **Human-centered computing** → **Interaction techniques**; **Virtual reality**; **Graphical user interfaces**; **Interaction design**; **Visualization systems and tools**.

KEYWORDS

spatiotemporal guidance, visual guidance, virtual reality, 360 degree video, cinematic virtual reality

ACM Reference Format:
Sean J. Liu, Rorik Henrikson, Tovi Grossman, Michael Glueck, and Mark Parent. 2023. RadarVR: Exploring Spatiotemporal Visual Guidance in Cinematic VR. In *The 36th Annual ACM Symposium on User Interface Software and Technology (UIST '23), October 29–November 01, 2023, San Francisco, CA, USA.* ACM, New York, NY, USA, 14 pages. https://doi.org/10.1145/3586183.3606734

1 INTRODUCTION

Today, cinematic virtual reality (CVR) is used in many film genres such as virtual tours, documentary, horror, music, and gaming. CVR uses omni-directional footage (i.e., 360° video) and offers a more immersive experience than conventional film [37]. However, viewers can only see a portion of the 360° scene at any time. As a result, they may miss important story events outside their field of view, such as a region of interest (ROI) behind them.

*Email: lsean@cs.stanford.edu. The author conducted part of this work during an internship at Reality Labs Research, Meta.
†Emails: rorik@henrikson.ca, {mglueck, mrkprnt}@meta.com
‡Email: tovi@dgp.toronto.edu

UIST '23, October 29–November 01, 2023, San Francisco, CA, USA

ACM ISBN 979-8-4007-0132-0/23/10…$15.00
https://doi.org/10.1145/3586183.3606734

To combat this issue, many visual guidance techniques have been proposed to redirect viewer attention [7, 16, 17, 33, 34, 51, 54]. While these techniques give viewers spatial guidance (*where* to look), they do not offer much temporal guidance (*when* to look). This can be problematic for 360° videos which contain ROIs that are short or have important entrances. The problem is exacerbated when viewing 360° videos in head-mounted displays, where it can take more time for a user to re-orient their viewpoint. For example, in the 360° music video *School of Rock* [52], students are scattered across a classroom and sing different parts of a song. During the finale, many phrases are very short, making it difficult to re-orient to the current singer before it was over. In some cases, it may be important to view the entrance of a ROI. In the horror 360° video *Help* [15], an alien breaks into a subway train while chasing the protagonists. While the alien is an important subject to see, its dramatic entrance (i.e., breaking into the train) adds to the horror effect and thus has a higher entertainment value at the beginning. In these examples, the *timing* of seeing ROIs is important.

This paper introduces *spatiotemporal guidance* for CVR, where both spatial and temporal information of ROIs is provided to the viewer. Temporal guidance indicates to viewers when an ROI is coming up and allows them to plan their head motion in advance of the next important story event. This helps viewers orient in both time and space to see short-duration ROIs and any associated entrance effects.

As a first step in exploring spatiotemporal guidance, we propose a new visual guidance widget, *RadarVR*. Our technique is inspired by blending two well known design metaphors: minimap navigation visualization [1, 13], and rhythm games [11, 23, 29]. Given pre-defined ROIs as input, RadarVR uses a radar visualization to represent a top-down view of the 360° scene and visualizes ROIs as moving wedges around the radar (Figure 1). The angular direction and radial distance of the wedges represent their location in space and time, respectively. As the video plays, the wedges move closer to the radar to indicate the passing of time. When a wedge reaches the radar, it indicates that the ROI is present and currently important. Using RadarVR, viewers can re-orient ahead of time to prepare for upcoming ROIs.

With RadarVR, viewers have both spatial and temporal information of ROIs in the video, which allows them to look ahead at future ROIs and to plan their head motion in advance. While it is possible to add temporal information to other types of spatial guidance techniques, we focused on the radar-inspired design to leverage familiar design metaphors of rhythm games when integrating temporal information. We propose design principles based on workshop feedback from early iterations of RadarVR.

In a user study, we compare two versions of RadarVR, *spatiotemporal* and *spatial-only*, to isolate any impact of temporal guidance and to investigate the trade-off between these two types of visual guidance. Our results show that the spatiotemporal version significantly helps users see more ROIs (6%), see the start of ROIs more frequently (81%), and view ROIs for longer durations (21%). Based on user feedback, we give an in-depth discussion of the trade-offs and subjective preferences between spatiotemporal and spatial-only guidance. We conclude with design implications for future work on spatiotemporal visual guidance techniques.

In summary, we make the following contributions in this work:

- The concept of spatiotemporal guidance in CVR, and its associated design principles;
- RadarVR, a new widget that offers spatiotemporal visual guidance;
- A user study exploring the trade-offs between spatiotemporal and spatial-only guidance, using two versions of RadarVR;
- Design implications for future work on spatiotemporal visual guidance techniques.

2 RELATED WORK

2.1 Gaze Guidance

Our work contributes to the space of gaze guidance techniques for cinematic virtual reality (CVR). Gaze guidance have been explored in many domains, including visual [7, 16, 17, 33, 34, 51, 54], audio [40, 51], and haptic [27, 28]. A taxonomy of existing techniques can be found in Rothe et al. [50]. These techniques explore various ways to provide spatial guidance (i.e., *where* to look); however, they do not offer much temporal guidance (i.e., *when* to look). As a first step in exploring temporal guidance, we focus on the visual domain to leverage familiar visual metaphors from popular games.

Many studies have investigated typical gaze patterns of images [56] and videos [38, 39, 53] in virtual environments. Some techniques change the virtual environment to adapt to the user's viewing behavior. For example, Pavel et al. [47] reorient 360° scenes at shot boundaries based on the user's viewpoint. Liu et al. [35] extend the time in a scene via video textures to wait for viewers to look in a salient direction. In contrast to approaches which change the virtual environment to adapt to the user, our technique focuses on guiding users to see ROIs without altering the existing 360° video.

Using Nielsen et al.'s taxonomy [45], existing visual guidance techniques can be categorized as implicit or explicit, as diagetic or non-diagetic, and as limiting or allowing interaction. Implicit cues provide subtle guidance [7, 16, 17, 49], whereas explicit cues are more overt [33, 34]. Diegetic cues are embedded within the narrative of the virtual environment [9, 30, 51, 54], whereas non-diegetic cues are external to the story [33, 34]. Techniques that limit interaction impose physical constraints on the viewer's movements (e.g., with a motorized swivel chair [21]), while those that allow interaction do not. Many experiments have explored the trade-offs among these techniques [10, 57, 59]. They fall on different parts of the Narrative Paradox [36] spectrum between author's control and viewer agency. Our technique falls under explicit, non-diegetic cues which do not limit interaction; RadarVR shows spatial and temporal information of ROIs overlaid on a given 360° video.

2.2 Spatial & Temporal Visualizations

Many visualization techniques have been proposed for spatial navigation. Halo [3], Wedge [22], and other arrow-based visualizations [8] provide spatial information of offscreen targets to help users efficiently localize them in 2D space. Other works explore visualizations for multi-scale navigation in 3D space [41, 42] and for out-of-view objects in VR and AR [18, 19, 61]. While these techniques are effective, they are not designed for CVR environments with spatially and temporally changing elements.

Table 1: Workshop Participants. Novice: tried VR a few times. Intermediate: fairly familiar. Expert: confident.

Participant	Rounds	Occupation	VR Experience
P1	1	Software Engineer	Novice
P2	1,2	Researcher	Intermediate
P3	1	Research Designer	Intermediate
P4	2	Software Developer	Expert
P5	2	Researcher	Intermediate

Several spatiotemporal visualizations have been proposed for 360° video navigation [31, 44, 55]. Vremiere [44] uses a Little Planet minimap (stereographic projection) with a timeline. Similarly, Route Tapestry [31] maps an extracted orthographic projection on a timeline for efficient navigation. Lilija et al. [32] allows viewers to interact with object trajectories over time via direct manipulation. While these visualizations are helpful for navigation, they are not designed for gaze guidance and display more information (such as non-ROIs) than necessary during playback. To minimize visual load, RadarVR uses simple visualizations to encode the spatial and temporal positions of ROIs.

Other works have also explored similar radar visualizations for 360° videos. Brillhard visualized points of interest in space and time as dots on circular rings [6]; however, their purpose was to analyze the orientation of scenes at cuts and did not explore real-time spatiotemporal visualizations for guiding viewer attention. Other works [5, 20, 50] have proposed radar visualizations to help viewers locate points of interest in 3D space; however, these techniques do not encode temporal information. Our visualization is designed to provide both spatial and temporal guidance for viewers.

2.3 Navigation in Games

Our work was inspired by two well-known design metaphors in games: mini-map visualizations for spatial navigation (Figure 1a), and temporal visualizations in rhythm games (Figure 1b). Examples of games with minimaps include Grand Theft Auto [13] and Call of Duty [1]. In these games, the top of the radar typically represents the front-facing direction, and the bottom of the radar represents the direction behind the user.

Popular rhythm games such as Dance-Dance Revolution [29], Guitar Hero [23], and Beat Saber [11] visualize time as distance from some reference position. In these games, the player's objective is to hit some moving targets by providing input at specific times (e.g., hitting a key). As time passes, all the targets move towards the reference position at the same speed. Beat Saber's 360 mode [12] also includes a spatial component by extending the range of moving targets to all directions. To leverage the familiar design metaphors from these games, RadarVR blends the spatial and temporal visualizations by encoding ROIs as moving wedges around a radar.

3 DESIGN PRINCIPLES & DERIVATION

We iterated on the design of RadarVR by running two rounds of internal workshops. Based on our goal of delivering spatiotemporal information and the workshop feedback, we derived the following design principles:

Figure 2: Early prototypes [2]. (a) Blocks representing ROIs were directly embedded in the 360° scene (top-down view). (b) Snap-to-periphery: out-of-view blocks snapped to the periphery. (c) Ring: the ring bands indicated out-of-view blocks.

D0. Leverage known design metaphors for user familiarity.
D1. Minimize cognitive load imposed on the viewer. Maximize focus on the video content.
D2. Keep the ROI visualization simple.
D3. Provide simple and explicit guidance cues for out-of-view ROIs.
D4. Show viewers a minimum amount of spatial and temporal information needed to plan their head motion.

Table 1 shows information of our workshop participants. In each round of workshops, we had each participant watch two videos and conducted open-ended interviews to collect their feedback. The videos were 2-min clips from professionally made 360° videos [2, 43].

We experimented with two prototypes, shown in Figure 2b, c. Both consisted of moving blocks on a ground plane, embedded directly in the 360° scene (Figure 2a). Each block represents an ROI in space and time, indicated by its direction and distance from the viewer. The block's distance represented the time until the ROI appeared, which was inspired from moving targets in rhythm games [11, 23, 29]. We experimented with different levels of guidance for out-of-view ROIs in our initial prototypes. In the first prototype (Figure 2b), we used an explicit directional cue: out-of-view blocks snapped to the viewer's headset periphery, and the thickness of the block showed how far the viewer had to rotate to see it. In the second prototype (Figure 2c), we used an implicit directional cue: each block came with a ring band visible in all directions to make viewers aware of out-of-view blocks.

First Round Feedback and Changes. All participants found the temporal aspect of our visualization intuitive and used it to reorient their viewpoint in advance without trouble. This provided motivation to maintain a final design that also leveraged existing design metaphors (**D0**).

All participants mentioned that it was cognitively demanding to monitor the stream of blocks, consume the video, and search for out-of-view blocks at the same time. P1 wished the blocks always appeared within their FOV, so they did not have to look down to monitor the blocks. Hence, to reduce cognitive load (**D1**), we attached the visualization in top-down view as an on-screen widget.

Initially, we labeled moving ROIs in our videos with more granularity. For example, in one video, we represented a moving elephant with a narrow sheared block instead of a single wide block. However, P1 thought the shape was too complex for the scene. P3 reported they interpreted the block to mean a transient important event and were afraid to look away for the duration of the block. Due to

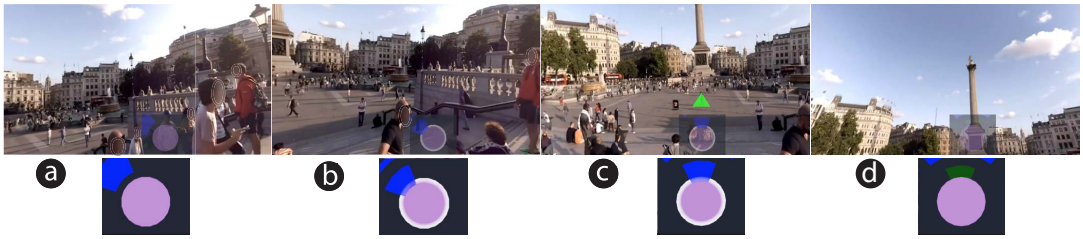

Figure 3: RadarVR represents ROIs as wedges around a radar. In this example [46], the statue atop the column is marked as an ROI when the narrator introduces it. (a) A wedge moves towards the radar, showing that an ROI is coming up. (b) The wedge hits the radar, indicating that the ROI is currently active. (c) An arrow provides vertical spatial guidance. (d) The wedge fades after the user sees the statue. Bottom row: zoomed-in versions of the widget (with increased opacity, brightness, and contrast for illustration purposes).

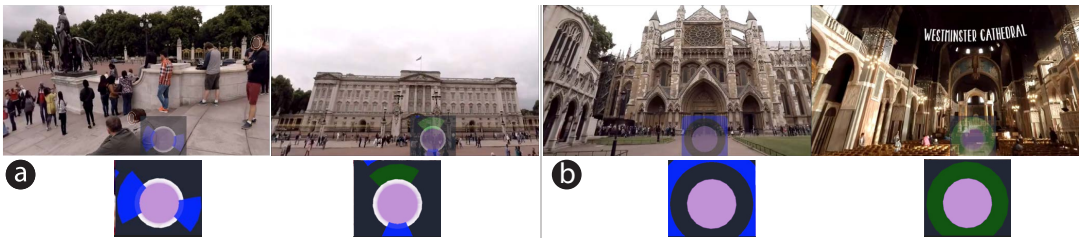

Figure 4: RadarVR can show various arrangements of ROIs [46]. (a) Multiple concurrent ROIs are represented as multiple wedges hitting the radar simultaneously. The wedges turn green and fade out as viewers turn to see them. (b) All-directional ROIs are used when all directions are equally important, e.g., the narrator describes the general interior decorations of the church. They are represented as donuts around the radar which turn green and fade after the viewer sees the scene. Bottom row: zoomed-in versions of the widget (with increased opacity, brightness, and contrast for illustration purposes).

these feedback, we learned to keep ROI visualizations simple (**D2**) to manage expectations and to lower cognitive demand.

In both prototypes, participants requested explicit and simpler cues for where to look and how much to turn to see out-of-view blocks (**D3**). In the snap-to-periphery prototype, although we encoded distance and the amount of rotation of out-of-view blocks in our visualization, we found that participants could not easily follow it. In the case of multiple blocks, it wasn't immediately clear which one they should look at first and how far they needed to turn to see it. We found that presenting less information at once makes it easier to follow (**D4**), so in our final design, we adopted a coarse-to-fine approach of spatial guidance.

Second Round Feedback and Changes. Based on the first round of workshop feedback, we discarded the ring and snap-to-periphery prototypes and added the top-down view of the visualization as an on-screen widget. We also adjusted the ROI visualizations to be simple wedges that disappear after viewers have seen the corresponding ROI.

The second round feedback confirmed the design principles from the previous round. For example, P5 suggested further simplifying ROI labels by merging adjacent separate blocks into one larger block (**D2**).

All participants had no trouble following the radar mini-map; P4 and P5 said they liked the temporal aspect for planning head motion. However, P5 said they did not know how long to keep looking in a direction after a wedge has disappeared. Hence, we adjusted the design to retain ROI duration info; instead of making the wedges completely disappear, we reduce their opacity once the ROI has been seen.

P2 commented that the wedges moved slowly, and monitoring and waiting for the wedges to arrive can be cognitively demanding (**D1**). We speculate a trade-off between look-ahead time and cognitive demand. A longer look-ahead time gives more heads-up time for viewers to react but may be mentally demanding to monitor (i.e., wedges are more compactly spaced on the map and move slower), and vice versa. Hence, to further lower cognitive load, we reduced the look-ahead time in the final design. We also removed the ground plane to remove redundant information (**D4**).

4 RADARVR: SPATIOTEMPORAL VISUAL GUIDANCE

We present RadarVR, a widget that provides spatiotemporal visual guidance in cinematic virtual reality (CVR). Here, we give an overview of RadarVR and explain our final design decisions. We discuss technical implementation details in Appendix A.

RadarVR takes as input a 360° video and a series of labeled ROIs. During video playback, RadarVR visualizes the ROIs around a radar overlaid over the video (Figure 3). In our system, a labeled ROI consists of a bounding box and a start and end time. Inspired by minimap radars [1, 13] and rhythm games [11, 23, 29], we visualize ROIs as wedges that move towards a circular radar (Figure 3a).

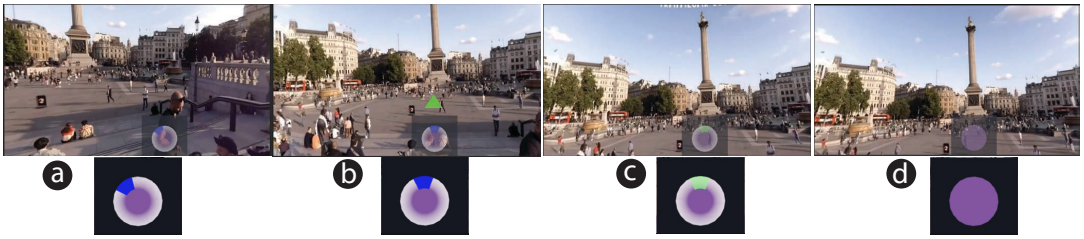

Figure 5: RadarVR (Spatial-only mode). (a) When an ROI is active, the widget provides spatial guidance by coloring the ROI direction blue. (b) Arrow gives vertical guidance. (c, d) Color fades out after viewer sees the ROI (i.e., statue atop the column). Bottom row: zoomed-in versions of the widget (with increased opacity and contrast for illustration purposes).

This design combines two familiar concepts for spatiotemporal visualization (**D0**). The angular direction of a wedge represents its spatial direction (in the yaw dimension), and its distance from the radar center represents its distance in time. As the video plays, the ROI wedges move towards the radar center at the same speed. When a wedge hits the circle, it indicates that the ROI is present and active, and the entire circle rim lights up to notify the viewer (Figure 3b).

The radar is attached to the bottom of the display and gives viewers a top-down overview of the 360° scene. This design compactly displays ROI information in one screen area and obviates the need to search for out-of-view ROIs, which could be cognitively demanding as we learned from the workshops (**D1**). To keep the ROI visualization simple, we use a simple wedge shape to represent each ROI (**D2**). The subtended angle of the wedge displays horizontal (yaw) direction of the ROI, and the radial thickness of the wedge represents its duration.

To show where other ROIs are relative to the viewer's current orientation and how far the viewer has to turn to see them (**D3**), the RadarVR visualization always aligns with the viewer's orientation, i.e., the top of the radar always points towards the viewer's front-facing direction, and the bottom of the radar points refers to the direction behind the viewer (similar to a compass).

To reduce the amount of spatial information displayed to the viewer (**D4**), RadarVR provides incremental spatial guidance. RadarVR initially only displays horizontal (yaw) spatial direction of ROIs on the radar. After the viewer orients to the correct horizontal direction, if the ROI is above or below the viewer's FOV, RadarVR then displays a green arrow to provide vertical guidance (Figure 3c). For typical 360° videos, however, objects of interest are on the ground plane, so an arrow is not needed. Finally, we limit the look-ahead time of ROIs to reduce the amount of temporal information shown (**D2 & D4**). From pilot testing, we found 10 seconds to be a good buffer time.

To reduce cognitive demand and keep the visualization simple (**D1** & **D2**), RadarVR reduces the visual saliency of wedges once an ROI is seen and viewers no longer need to anticipate its arrival. Specifically, once the viewer sees an active ROI, RadarVR marks the corresponding wedge as "hit" and provides visual confirmation by turning its color to green and making it very faint (i.e., low opacity).

Some scenes may have more than one ROI at the same time, in which case RadarVR shows multiple wedges hitting the radar simultaneously (Figure 4a). In other scenes, all directions could be equally important (e.g., a narrator might generally describe the interior decorations of a building). To take these cases into account, RadarVR also accepts all-directional ROI labels and visualizes them as donuts around the radar (Figure 4b).

5 USER STUDY

We created RadarVR as a first step in exploring spatiotemporal visual guidance. While there are many existing spatial guidance widgets, our goal was not to compare our design to those techniques, as they do not provide temporal information. Rather, our goal is to explore the concept of spatiotemporal guidance and to isolate the effects of adding temporal information. To this end, we conducted a user study using RadarVR to compare *spatiotemporal* and *spatial-only* guidance. The purpose of our study is to reveal the potential benefits of temporal guidance, which could inform the designs of adding temporal information to other spatial guidance techniques.

5.1 Research Questions

With spatial-only guidance as the baseline method, we explore the following research questions:

RQ1. How does spatiotemporal guidance affect the **percentage** and **duration** of **ROIs** seen?
RQ2. How does spatiotemporal guidance affect **head motion planning** (i.e., do viewers orient their view in advance of an upcoming ROI)? How are these effects impacted by video characteristics?
RQ3. How does spatiotemporal guidance affect **cognitive load** and feelings of **FOMO** (fear-of-missing-out)?
RQ4. What is the **perceived usability** of RadarVR's spatiotemporal guidance?
RQ5. What are viewers' **subjective preferences** between spatiotemporal guidance and spatial-only guidance?

5.2 RadarVR: Spatial-only Mode

To compare spatiotemporal and spatial-only guidance, we created a *spatial-only* version of RadarVR for our study (Figure 5). While there are other existing spatial-only guidance widgets [33, 34], we chose to compare with a spatial-only version of RadarVR in order to minimize the difference between the two versions and isolate the effects of temporal guidance. We do not claim or seek to show that RadarVR's spatial-only mode is better than existing spatial-only techniques; rather, it was chosen as a baseline as it provides us with the smallest differential to the RadarVR widget. Future work could

Table 2: User Study Experiment Videos. The 360° videos represent a range of genres and spatial and temporal distributions of ROIs. For each video, we list statistics about the duration of ROIs, the lateral angle (yaw) between time-adjacent ROIs, as well as the time between time-adjacent ROIs.

Name	Genre	Total # of ROIs (1st + 2nd halves)	Duration (s) μ, σ, [Min, Max]	Angle $\in [0°, 180°]$ μ, σ, [Min, Max]	Time (s) μ, σ, [Min, Max]
France [2]	Virtual Tour	23 (12 + 11)	8.2 5.2 [2.5, 20]	54 56 [0, 167]	15 10 [2.5, 43.5]
Elephants [43]	Documentary	23 (12 + 11)	10 4.9 [3, 21]	89 59 [0, 177]	14 10 [0, 39]
Help [15]	Horror	43 (27 + 16)	3.5 4.0 [1, 25]	111 56 [7, 180]	6.6 5.6 [1.5, 31.5]
School of Rock [52]	Music	51 (20 + 31)	4.2 3.8 [1, 23]	40 49 [0, 169]	5.8 4.2 [1.3, 24]
Pokémon [25]	Game	38 (21 + 17)	21 18 [1, 60]	85 52 [4, 169]	5.8 7.9 [0, 33.5]

compare the RadarVR to spatiotemporal adaptations of existing guidance widgets.

RadarVR's spatial-only mode has the same features as the spatiotemporal one, except it does not show any look-ahead time or duration of ROIs. When an ROI is active, the spatial-only mode provides spatial guidance by coloring the ROI direction blue on the radar (Figure 5a). Once the viewer sees the active ROI, the colored portion turns green, but unlike the spatiotemporal version, it completely fades out and does not show any duration info (Figure 5c, d). To make sure viewers see the colored portion during an active ROI, we make the radar rim thicker in the spatial-only version. Otherwise, the features between the two versions remain the same.

5.3 Experiment Design

To investigate the research questions posed in Section 5.1, we ran a 2 x 5 within-subjects experiment using two widget conditions, spatial-only (**S**) and spatiotemporal (**ST**), and five professionally made videos. We selected the videos to represent a range of film genres and ROI distributions (Table 2). We give video descriptions and ROI labeling details in Appendix B. To prevent participants from watching the same video twice (under the two conditions), we cut each of the five videos roughly in half and assign them into two groups. The first half of all videos are assigned to the first group, and the second half of all videos are assigned to the second group. This created ten $2-3$ minute videos in total.

During the study, participants watched the first group of videos during their first assigned condition (S or ST), and then watched the second group of videos for their second assigned condition (ST or S). The two widget conditions were counterbalanced across participants, and the video orders were randomized within each condition.

Before each condition, the participants went through a tutorial that explained the widget features and experienced a 1.5-minute practice demo to become familiarized with the widget. Because our goal is to investigate the widgets' performance of visual guidance, participants' task was to see as many important regions as possible with the support of the widgets.

5.4 Experiment Measures

We used the following data measures to address the research questions in Section 5.1.

Percentage and Duration of ROIs Seen & Head Motion Planning. During the study, we collected head tracking data of participants. Based on the data, we computed the percentage of ROIs seen and the duration percentage of ROIs seen. For head motion planning, we computed the percentage of ROIs where the participant saw their initial onset.

Cognitive Load & FOMO. After watching all videos in a condition, we asked participants to fill out a NASA-TLX questionnaire [24] to measure task load as a proxy for cognitive load. We also asked them to rate their level of FOMO (on a 7-point Likert scale).

Perceived Usability. After both conditions, participants filled out a comparison survey to record the perceived usability of each widget for completing their tasks. Participants rated the following statements (on a 7-point Likert scale):

(1) *[S or ST] successfully helped me see important regions in the videos*;
(2) *[S or ST] helped me plan my head motion to see the start of important regions*;
(3) *It was easy to use [S or ST] for watching videos.*

Subjective Preference. In the comparison survey, we also asked participants to rate their overall preference between the spatiotemporal and spatial-only versions (on a 7-point scale) and what they liked and disliked about each one. In addition, we collected their preference between the two versions for each individual video (Table 2). Finally, we conducted brief, open-ended interviews to collect general feedback.

6 RESULTS

6.1 Participants

We recruited 16 external participants (9 male, 6 female, 1 preferred not to say their gender). Their ages ranged from 19 to 28 ($\mu = 23.6, \sigma = 2.3$). Out of 16 participants, 14 were undergraduate or graduate students, one was an IT operations staff, and one was a UI/UX designer. Our participant pool had a range of VR and 3D gaming experiences. Out of 16 participants, 5 had no prior VR experience, 6 have tried VR 1-5 times, 4 have tried VR 5-20 times, and 1 has tried VR more than 20 times. Most participants who have experienced VR use it for gaming. Regarding how often they played 3D games, 8 answered never or rarely, 2 answered sometimes, and 6 answered often or always.

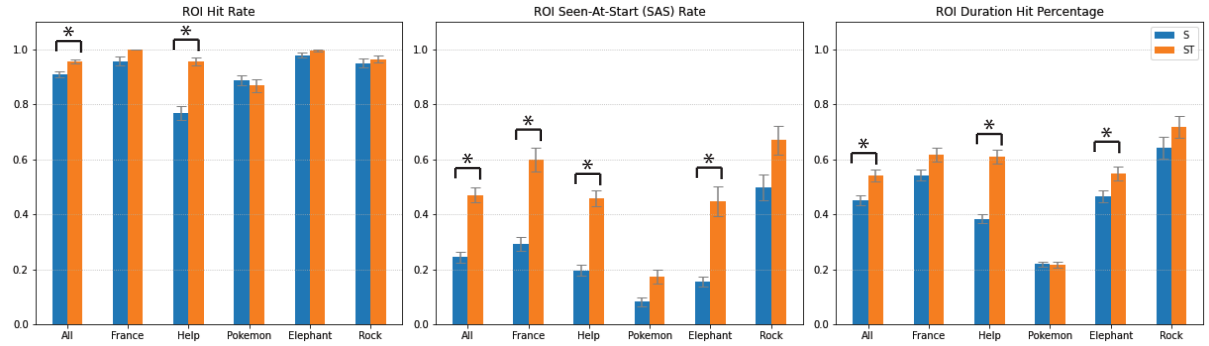

Figure 6: Quantitative Results of Viewing Behavior. (*) indicates a significant main effect or simple main effect with Bonferroni correction [4]. Error bars denote standard error of mean.

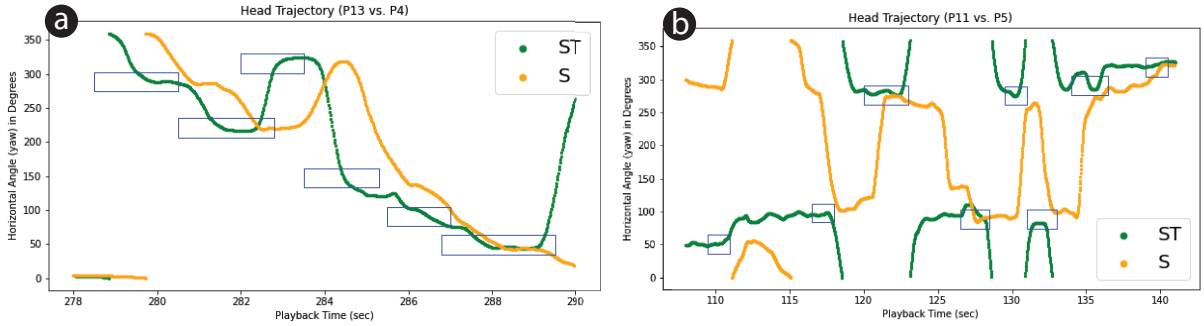

Figure 7: Example Head Trajectories of ST (green) vs. S (yellow). Horizontal axis: video playback time. Vertical axis: head orientation in the horizontal (yaw) direction. ROIs are represented as rectangles on the graphs. (a) Head trajectories of P13 (ST) and P4 (S) during the finale of the *Rock* video [52], where there was a quick sequence of short ROIs. Despite the short ROI durations, the viewer in ST was able to hit all the ROIs, whereas the viewer in S only hit half of them. (b) Head trajectories of P11 (ST) and P5 (S) during the chase in *Help* [15]. The viewer in ST saw all the ROIs from their initial onset, whereas the viewer in S did not see the entrance of any ROIs.

6.2 Viewing Behavior

We analyzed participants' head tracking data to measure viewing behavior differences. We used two-way repeated measures ANOVA tests to evaluate the effect of widget condition and video on the following three data measures. In all cases, there were significant interaction effects between widget and video, so we ran follow-up tests to check the simple main effect of widget for each video, with adjusted p-values using Bonferroni correction [4].

6.2.1 ROI Hit Rate. For each video, we computed the percentage of ROIs seen under both conditions. Given an ROI, if the viewer saw the correct spatial region any time between the ROI's start and end time, we considered it a "hit." Compared to the **S** condition, the ROI hit rate across all videos increased by 6% in **ST**. There were main effects of widget ($F(1, 15) = 10.418, p = 0.006, \eta^2 = 0.132$) and video ($F(2.65, 39.81) = 39.301, p < 0.001, \eta^2 = 0.415$), as well as an interaction effect between video and widget on ROI hit rate ($F(1.95, 29.26) = 17.184, p < 0.001, \eta^2 = 0.249$). The simple main effect of widget is significant for *Help* ($p_{adj} < 0.001, \eta^2 = 0.571$) but not the other videos.

6.2.2 ROI Seen-At-Start (SAS) Rate. For each video, we computed the percentage of ROIs where the participant saw the correct region at the start time. There were main effects of widget ($F(1, 15) = 27.672, p < 0.001, \eta^2 = 0.4$) and video ($F(4, 60) = 116.739, p < 0.001, \eta^2 = 0.551$), as well as an interaction effect between video and widget on ROI SAS rate ($F(2.27, 33.98) = 3.845, p = 0.027, \eta^2 = 0.081$). The simple main effect of widget is significant for *Elephant* ($p_{adj} = 0.002, \eta^2 = 0.474$), *France* ($p_{adj} < 0.001, \eta^2 = 0.567$), and *Help* ($p_{adj} < 0.001, \eta^2 = 0.654$). Compared to the **S** condition, the ROI SAS rate across all videos increased by 81% in **ST**. This suggests that the temporal feedback was significantly effective in helping viewers plan their head motion.

6.2.3 ROI Duration Hit Percentage. For each ROI, we computed the percentage of frames where there was an ROI hit. On average, the program logs a frame once every 0.014 seconds. Compared to the **S** condition, the ROI duration percentage across all videos increased by 21% in **ST**. There were main effects of widget ($F(1, 15) = 25.883, p < 0.001, \eta^2 = 0.177$) and video ($F(4, 60) = 206.378, p < 0.001, \eta^2 = 0.711$), as well as an interaction effect between video and widget on ROI duration percentage ($F(1.94, 29.11) = 5.193, p =$

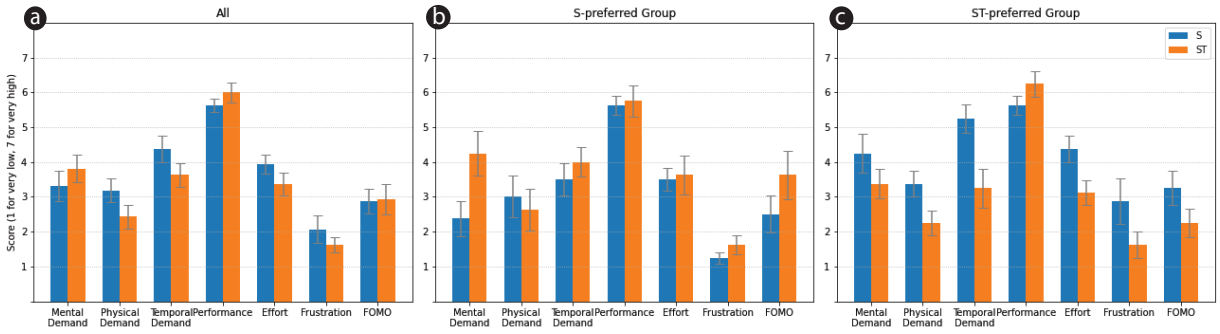

Figure 8: NASA-TLX [24] & FOMO scores. Error bars denote standard error of mean. (a) Overall, a two-way mixed ANOVA analysis shows no significant main effect of widget (S: spatial-only; ST: spatiotemporal) on any of the scales. (b, c) However, there is a significant interaction effect between widget and preference. When we group participants based on their preference (S vs. ST), we notice some interesting patterns.

0.012, $\eta^2 = 0.124$). The simple main effect of widget is significant for *Elephant* ($p_{adj} = 0.045, \eta^2 = 0.184$) and *Help* ($p_{adj} < 0.001, \eta^2 = 0.647$).

Taken together, these quantitative results show promise for spatiotemporal visual guidance. With temporal feedback, viewers overall see more ROIs, see the start of ROIs more frequently, and view ROIs for longer durations (Figure 6). Generally, we see an average increase from **S** to **ST** in all three measures. The increases are significant for *Help* in all three measures. This is likely because *Help*'s time-adjacent ROIs are on average spatially far apart (Table 2), so viewers need more time to turn and see them (e.g., the alien and protagonists are 180° opposite from each other). As a result, the look-ahead time of **ST** proves particularly effective in this video compared to the other ones. We also see some significant simple main effects of widget in *France* and *Elephant*, where the tour and documentary narrations offer less heads-up of future ROIs. In these cases, **ST** is also particularly effective because it offers temporal guidance. In contrast, *Pokémon* does not show an increase in ROI hit rate or duration hit percentage. One possible reason may be due to the absence of a narrative in the video; viewers are instructed to actively search for Pokémons and thus hit about the same number and duration of ROIs in either condition. Future studies would be needed to investigate these variation effects across a larger sample of videos.

Figure 7 shows examples of head trajectories in the **ST** and **S** conditions. In the Rock finale (Figure 7a), different singers sang short phrases across the classroom, resulting in a sequence of short ROIs. The viewer in **ST** was able to see all of the singers (marked as ROIs), whereas the viewer in **S** only saw half of them. Figure 7b shows the sequence of ROIs during the alien chase, where the alien and protagonists were 180° opposite from each other. The viewer in **ST** was able to plan their head motion and see all the ROIs at their start time, whereas the viewer in **S** did not see the beginning of any ROIs.

6.2.4 Angular Distance Travelled. To understand how temporal guidance might affect scene exploration, we also computed the horizontal (yaw) angular distance traveled per participant for each clip. Each clip was viewed by all 16 participants, either under the **S** or **ST** condition. For each clip, we compared the average angular distance traveled under each condition but did not find a significant difference using the Wilcoxon signed-rank test [60] ($p = 0.13; \mu_{st} = 3874°, \sigma_{st} = 875°, \mu_s = 4042°, \sigma_s = 793°$). The average angular distance traveled is slightly lower for **ST**. This is likely because viewers look around more in **S** in the absence of a heads-up for the next ROI, whereas viewers have more planned, focused head motions with temporal guidance.

6.3 Usability

6.3.1 Cognitive Load. Participants filled out the NASA-TLX questionnaire [24] after each widget condition. Results are shown in Figure 8a. Our analysis shows no significant main effect of the conditions on any of the dimensions.

6.3.2 FOMO. We asked users to rate their level of FOMO (fear-of-missing-out) for each condition (1 for very low, 7 for very high). Our results show no significant main effect of widget condition on FOMO (Figure 8a). From their free response answers, we found that the type of FOMO experienced may be different between the **ST** and **S** versions. The FOMO in **S** stems mainly from not knowing upcoming events in the video and possibly missing the content itself, whereas the FOMO experienced in **ST** could stem from two areas: 1) being aware of upcoming targets and afraid of missing them, 2) missing non-ROI regions in the video while looking at ROIs.

6.3.3 Perceived Usability. Participants rated statements (on a 7-point Likert scale) about perceived usability for **ST** and **S**. We used the Wilcoxon signed-rank test [60] on paired scores of **S** and **ST** to evaluate significance. For *planning head motion*, **ST** scores were significantly higher than **S** ($p < 0.001; \mu_{st} = 6.4, \sigma_{st} = 0.9; \mu_s = 2.5, \sigma_s = 1.7, r = 0.606$), which matches the objective increase in ROI SAS rate. For *ease of use*, there was no significant difference between **ST** and **S** ($p = 0.75; \mu_{st} = 5.2, \sigma_{st} = 1.7; \mu_s = 5.3, \sigma_s = 1.8$). For *seeing important regions*, **ST** and **S** also had no statistical difference ($p = 0.16; \mu_{st} = 6.1, \sigma_{st} = 1.1; \mu_s = 5.6, \sigma_s = 1.0$). Although participants objectively saw more ROIs in the **ST** condition, there was no significant increase in viewers' perception of ROI hits. This may be because viewers are less aware of the ROIs they missed in

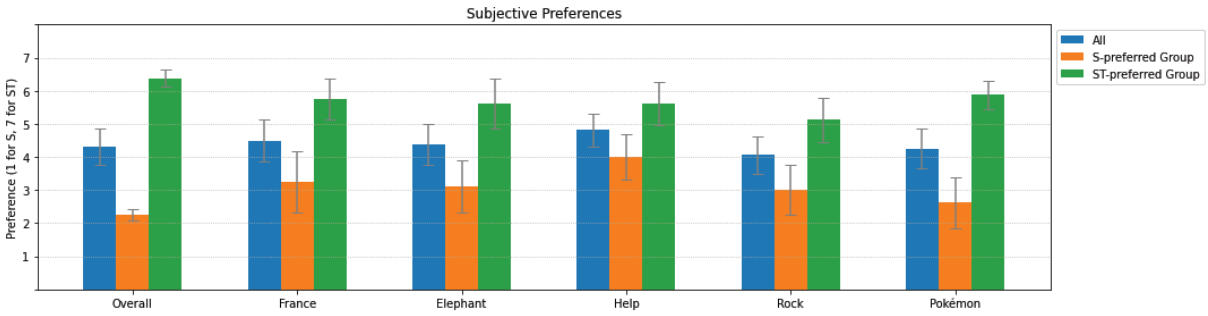

Figure 9: Subjective preferences between the spatial-only (S) and spatiotemporal versions (ST). 1 for S, 4 for no preference, and 7 for ST. Error bars denote standard error of mean. When we divide the samples into two groups (overall prefer S vs. overall prefer ST), we see that the difference between the group averages is smallest for the Help video (non-concurrent short ROIs), and largest for the Pokemon video (multiple concurrent ROIs).

the **S** condition (**P3:** "*it is a lot easier to not notice that an action even came and went with the spatial-only widget.*"). Due to the look-ahead time in the **ST** condition, viewers get more heads-up notice of ROIs and are thus likely more aware of missing them.

6.3.4 Spatial Mapping. Most participants (i.e., 14 out of 16) had no issue with the spatial mapping of RadarVR (i.e., top indicates front, bottom indicates back). One participant was slightly confused at first but became familiarized with the mapping after one video and called the visualization "intuitive." The other participant only mentioned difficulty deciding whether to turn counterclockwise or clockwise in the *Help* video, where the ROIs were 180° opposite from each other and occurred in quick succession. We believe the spatial mapping of RadarVR is easy to understand because it leverages familiar design metaphors from 2D map navigation, such as Google Maps [14].

6.4 Subjective Preferences

6.4.1 Overall & Video Preferences. Participants rated their preference between the *spatial-only* (**S**) and *spatiotemporal* (**ST**) widgets on a 7-point scale (1 for **S**, 4 for no preference, and 7 for **ST**). Participants reported an overall preference between the two versions, and then reported preferences for individual videos. Results show that overall preferences are very divided between the **ST** and **S** versions. Half of the participants (i.e., 8 out of 16) overall preferred **ST**, and the other half (i.e., 8 out of 16) overall preferred **S**. When we divide the samples into two groups (overall prefer **S** vs. overall prefer **ST**), we see the preference scores within each group are concentrated (Figure 9).

Preferences for individual videos are also shown in Figure 9. The difference between the group averages is smallest for *Help* [15] (which has many non-concurrent, short ROIs). This suggests that participants generally found **ST** helpful for capturing non-concurrent, short ROIs in fast-paced videos. The difference is largest for *Pokémon* [25], which has multiple concurrent ROIs. This suggests that task difficulty (i.e., hitting multiple concurrent ROIs) may be a driving factor for individual differences in preference.

6.4.2 Likes & Dislikes. Quantitative results suggest that individual differences in preference may be due to task difficulty (Section 6.4.1). Based on the free response and interview feedback, we found that individual preferences may also depend on their goals of watching 360° videos. In general, participants who preferred **ST** valued seeing important content, and those who preferred **S** valued freedom to explore videos at their own pace.

Participants who preferred **ST** stated they liked having a preview of upcoming targets and liked that **ST** gave them buffer time. **P5**: "*I preferred to have a preview of where I should look in advance rather than having the important points appear all of a sudden.*" **P2**: "*I prefer spatiotemporal because it simply gives way more reaction time.*"

Participants who preferred **ST** felt less pressure of missing important content and was able to plan their head motion. **P1**: "*The spatiotemporal one gives me enough buffer to be mentally ready to switch my attention, and I don't feel pressured.*" **P9**: "*Knowing beforehand...which region to focus on give me more time to act. Knowing the duration of the region of interest also let[s] me plan my focus. In short [the] more information the better.*"

Participants who preferred **S** said it gave them more freedom to look around, felt less stressed to hit all the targets, and liked that the **S** widget was simpler and required less attention. **P3**: "*[T]he spatial-only widget just allows for a more natural viewing experience. I think the requirement to look in a direction ahead of time in preparation of viewing an action deters away from enjoying a scene in the spatiotemporal case.*" **P7**: "*Spatiotemporal imposes more pressure on me to look at the blocks... The spatial-only felt more like a suggestion rather than instructions so I didn't feel bad if I missed something.*" **P4**: "*I'd prefer [spatial-only] because it points me to what I need to see when I need to see. There's less widget movement so I'm not distracted from the videos themselves.*"

As for individual videos, many participants stated they preferred **ST** for videos with short or multiple concurrent ROIs, because it allowed them to plan ahead. Even though the ROI hit rate and duration percentage did not objectively increase in the Pokémon video (Figure 6), many participants liked having temporal information for planning. **P2**: "*when there are many moving objects, or objects only existed for a short span of time, the extra temporal info is extremely helpful for viewer to observe important objects in time.*"

Participants also thought **ST** was useful for catching targets, and preferred it for videos where there are large spatial spreads between

ROIs so they do not need to turn suddenly. Some participants reported they preferred **S** for slow-paced and simpler videos because there is less need for visual guidance.

Other commonly liked features of **ST** include the display of ROI duration, which helped them prioritize what to look first. **P9**: "*Knowing the duration of the region of interest also let me plan my focus.*" A few participants liked how **S** notified them of ROIs as they occurred. However, they also mentioned it was hard to capture multiple ROIs or short ROIs.

6.4.3 Preference & Cognitive Load. Based on participant feedback, we hypothesize that mental demand may be a driving factor for preferring **S** over **ST**. Viewers who find **ST** mentally demanding may not enjoy it, while viewers who do not find **ST** mentally demanding may be able to use it with more success and ease.

To test this hypothesis, we ran a two-way mixed ANOVA to evaluate the effect of widget condition (within-subject) and preference (between-subject) for each NASA-TLX dimension. There were no significant main effects of widget or preference for any of the dimensions. However, there were significant two-way interactions between widget and preference for mental demand ($F(1,14) = 10.654, p = 0.006, \eta^2 = 0.189$), temporal demand ($F(1,14) = 6.731, p = 0.021, \eta^2 = 0.203$), and frustration ($F(1,14) = 6.326, p = 0.025, \eta^2 = 0.122$). We then checked whether the simple main effect of widget was significant for each preference group. We did not find any significant simple main effect after applying Bonferroni adjustment [4]. The effect of widget on mental demand in the **S**-preferred group was $F(1,7) = 7.915, p = 0.026, p_{adj} = 0.052$. The effect of widget on temporal demand in the **ST**-preferred group was $F(1,7) = 5.895, p = 0.046, p_{adj} = 0.092$.

Despite the lack of significance, the trends suggest that those who prefer **S** tend to find **ST** more mentally demanding (Figure 8b), and those who prefer **ST** tend to find **ST** less temporally demanding (Figure 8c). There may be a learning effect at play; 5 of the 6 participants who have significant 3D gaming experience prefer **ST**, so those who are more familiar with minimap navigation may enjoy it more. We leave up to future work to explore these effects further.

7 CONTENT CREATOR FEEDBACK

To understand what content creators thought of RadarVR, we conducted informal interviews with two professional CVR content creators and summarize their feedback here. The interviews were conducted over video call and lasted about one hour each. During the interview, we first gave an overview of the problem and asked them about their current practices and solutions. We then introduced the RadarVR visualization and showed examples of the widget on a few 360° videos.

Current Practices. Both creators agreed that viewers missing short events or the onset of events was a common problem for them. When first making 360° videos, one creator tried to use the whole 360° stage but soon realized that viewers often looked in the wrong direction. Their main solution was to modify their scripts to limit ROIs to the same spatial direction and to avoid quick or simultaneous events, which "limit[s] [their] storytelling" (**C1**). In some cases, they use techniques such as arrows, highlighting, blurring, text, or audio cues to direct viewer attention.

RadarVR Visualization. Both creators overall gave very positive feedback and were eager to use RadarVR in their storytelling. **C1**: "*[This is] a neat and clever way of solving the challenge...[RadarVR] can increase the complexity of storytelling...[and uses] a familiar concept / interface for anticipating events.*" **C2**: "*It's a good support [tool]....to guide the users to a better experience, and of course, as a storyteller, to tell the story as I imagined...I would definitely use it.*"

When asked about the disadvantages of RadarVR, one creator mentioned that RadarVR may eliminate elements of surprise for viewers, which may not be desirable if the filmmaker's goal is to surprise the viewer. As one solution, they suggested selectively turning RadarVR on and off. This suggests an interesting design space for using RadarVR in storytelling (e.g., when to show or hide an ROI given the type of story and desired effect), which future work can explore.

Motivated by our design goals and study results, we also asked creators about the potential visual disruption of using the widget. One creator agreed it is important to avoid breaking immersion, but both creators said they were not too concerned about it and believed that the widget can be blended very well into the background. **C1**: "*I'm not worried about it.*" **C2**: "*I'm sure [the blending] can be done very, very well.*" As examples, they suggested techniques such as fading out RadarVR when not needed and changing its brightness and transparency.

Usage of RadarVR. Both creators said they would use RadarVR in their filmmaking because it would allow them to take advantage of the full 360° stage to tell stories rather than being limited to a spatially concentrated region. In addition, they believe it is a far worse user experience to miss ROIs. **C2**: "*Missing a moment of the scene is much worse than having a little map in front of you.*"

We also asked their opinions on who should have control over enabling and disabling RadarVR (e.g., filmmaker vs. viewer). Both creators believed that the filmmaker should have primary control for better narrative control, with the viewer having the secondary choice to opt out (similar to skipping cutscenes in games). **C2**: "*I would encourage it to be there as much as possible...I don't want important part[s] of the story to be missed.*"

Narrative Paradox. Finally, the creators discussed their views on how RadarVR plays a role in the Narrative Paradox [36], i.e., the trade-off between author's narrative control and viewer agency. **C1** did not perceive it as a straightforward trade-off and believed RadarVR gave *both* the author more narrative control (i.e., use the full 360° stage while ensuring viewers see ROIs) and also gave the viewer more freedom (i.e., viewers know when and where ROIs are and can choose to view them). **C1**: "*If you don't know [what the options are], you can't choose.*" The other creator believed RadarVR gave the author more "passive control" by increasing the likelihood that viewers see ROIs without limiting viewers' freedom.

8 DISCUSSION

Here, we summarize main findings from our study, which compared two versions of RadarVR: *spatiotemporal* (**ST**) and *spatial-only* (**S**). Based on our findings, we discuss design implications for spatiotemporal visual guidance, which could inform the designs of adding temporal feedback to other spatial guidance techniques.

Table 3: Summary of design recommendations for spatiotemporal visual guidance, along with relevant findings from the study.

Design Recommendation	Relevant Study Findings
Use temporal guidance for fast-paced or short ROIs, where the onset and/or duration of the event is important	6.2
Use temporal feedback to build anticipation	6.3.2
Omit temporal feedback if the goal is to let viewers explore the scene at their own pace	6.4.2
Match ROI labels to what viewers deem to be important (to reduce FOMO)	6.3.2
Make temporal feedback optional to accommodate individual preferences	6.4.2
Offer different levels of temporal details to accommodate individual preferences	6.4.2
Minimize mental demand while providing sufficient temporal guidance	6.4.3

8.1 Main Findings

The main findings from our study are summarized below:

ST is more effective than S at visual guidance. Our analysis shows that **ST** helps viewers see more ROIs than **S**. The addition of temporal guidance successfully helps viewers plan their head motion with respect to the timing of ROIs, resulting in more frequent views, more timely views, and longer views. While viewers perceive that **ST** is more effective than **S** in planning head motion, they do not perceive a significant increase in ROI hits. This is possibly because viewers are less aware of the ROIs missed in **S**, whereas viewers get more notice of upcoming ROIs in **ST** and are thus more aware of the ones they missed.

No significant difference in cognitive load between ST and S. Our study found no significant difference in cognitive load between **ST** and **S** (as measured by NASA-TLX [24]) and no significant difference in perceived ease of use.

Individual preference between ST and S may vary. Our study shows that subjective preference may vary among individuals. One possible reason is due to differences in mental demand. Although there was no significant difference in cognitive load between **ST** and **S**, we found significant two-way interactions between widget and preference for mental demand, temporal demand, and frustration. While we did not find any significant simple main effects after Bonferroni correction, viewers who prefer **ST** on average found it less mentally demanding and less temporally demanding. They also reported lower levels of frustration. This suggests that mental demand is an important design consideration for future work in temporal guidance. It also suggests there may be a learning effect at play; viewers who find **ST** mentally demanding may not enjoy it, but viewers who do not find it mentally demanding enjoy it more and use it with more ease and success. Another possibility for difference in individual preference may be the viewer's goal. In particular, our qualitative feedback suggests that individuals who preferred **ST** over **S** valued seeing important content, whereas those who preferred **S** over **ST** valued the freedom to explore videos at their own pace.

No significant difference in FOMO between ST and S. Our study found no significant difference between levels of FOMO (fears of missing out) between **ST** and **S**. Qualitative feedback suggests this may be because the type of FOMO participants experience is different under each condition. The FOMO in **S** stems from not knowing upcoming events in the video and possibly missing the content itself, whereas the FOMO experienced in **ST** may stem from two areas: 1) anticipating upcoming ROIs and being afraid of missing them, 2) missing non-ROI regions in the video while looking at ROIs.

8.2 Design Implications

Our results reveal many design implications for applying temporal guidance to storytelling in CVR, summarized in Table 3. From the study, we learned that the addition of temporal feedback makes visual guidance more effective. As such, we suggest story designers use temporal guidance for important content that viewers should not miss, as well as events where the beginning is important or events where viewers should watch for an extended period of time. Temporal feedback is particularly useful when there is a need for head motion planning, such as in fast-paced scenes or scenes with short ROIs. From qualitative feedback, we learned that the temporal feedback may help build anticipation and make viewers feel more compelled to hit upcoming ROIs. As such, we recommend story designers use temporal feedback for story events where they want to build up anticipation and suspense.

While temporal guidance is effective at visual guidance, we recommend storytellers use temporal feedback selectively rather than all the time. Qualitative feedback suggests that when there is no temporal feedback, viewers tend to focus on the scene more and also feel more freedom to explore. Hence, we recommend omitting temporal feedback for less important content or in scenes where the director's intention is to have viewers explore the scene at their own pace. In addition, some viewers reported feeling FOMO for regions they do not see while following the labeled ROIs. This suggests a tradeoff in labeling ROIs; labeling an ROI makes it more likely that viewers see it, but it may also induce FOMO for regions outside of the marked ROI. To provide a positive experience, we recommend creating labels which match what viewers deem as important. Future work could investigate best storytelling practices for showing temporal feedback and blending spatial-only and spatiotemporal guidance.

Because individual preferences may vary, we recommend making temporal feedback an optional feature. That way, viewers who prefer to plan their head motion and see important events may enable the feature, while viewers who prefer to explore scenes at their own pace may disable it. By the same token, designers can also experiment with offering different levels of temporal details. That way, viewers who prefer a lot of guidance can choose to see all important regions, and viewers who prefer a small amount of guidance (and more freedom) can choose to see only a few most important regions.

Finally, we recommend minimizing mental load in future design iterations of spatiotemporal guidance. Although there was no overall significant difference in cognitive load between RadarVR's temporal and non-temporal modes, our analysis suggests that viewers may dislike the temporal feedback if it is mentally demanding for them to use. Future iterations could further improve the design to lower cognitive demand while providing sufficient temporal guidance (e.g., non-continuous temporal visualization, shorter or adaptive look-ahead time).

9 LIMITATIONS & FUTURE WORK

RadarVR takes pre-defined ROIs as input and visualizes them during playback. Future work could investigate automatic labeling or new interfaces to support labeling.

Because the goal of our study was to investigate RadarVR's performance of visual guidance and to evaluate the effect of temporal feedback, participants' task was to see as many important regions as possible. As such, the study results may be related to the specific goals of the study, and further studies could evaluate RadarVR with other tasks (e.g., watch videos naturally and can choose to ignore the widget) and quantitatively measure other aspects (e.g., immersion, distraction).

In RadarVR, we added two types of temporal information: when an ROI will arrive in time (look-ahead time), and how long an ROI will last (duration). However, our interview feedback suggests that viewers may value these two types of information differently. While some participants did not like the anticipation of seeing upcoming ROIs (and felt more pressure), many of them said they liked knowing how long a scene will last, even for slow-paced scenes. Future work could investigate these two dimensions separately and explore their relative importance in storytelling.

RadarVR presents many interesting opportunities for filmmaking in CVR. As shown in the feedback from content creators, directors often spatially concentrate ROIs to ensure viewers see important events, at the expense of under-utilizing the rich spatial potential of the view-sphere. RadarVR may enable directors to direct scenes with more spatially spread-out or temporally concentrated ROIs, with its effective spatiotemporal visual guidance. As such, there is a rich design space for choosing when to show temporal feedback and when to hide it. Future work could explore this design space in a range of scenes and video types. Directors can explore new ways to design storytelling experiences with this widget and with spatiotemporal feedback.

Future work can enhance RadarVR's UI design to blend more seamlessly with the scene and maintain immersion, i.e., altering position, size, and visibility. To control for these variables in the study and isolate the effect of temporal guidance, we kept the widget centered and visible at all times, and we showed the vertical arrow above the widget to more easily grab viewer's attention near the center of their FOV. While we designed the ROI visualizations within the widget to reduce cognitive load, future work can experiment with alternative UI designs to further reduce visual saliency of the widget (e.g., fading out the widget when current and upcoming ROIs are within the viewer's FOV, placing RadarVR in a corner).

Future work can explore ways to add temporal guidance to other types of spatial guidance techniques. While we focused on the visual domain as a first step in spatiotemporal guidance, it would be interesting to explore spatiotemporal guidance using other modalities, such as audio and haptics.

10 CONCLUSION

We presented a new widget, *RadarVR*, as a first step in exploring spatiotemporal guidance in CVR. We proposed a set of design principles for spatiotemporal visual guidance and conducted a study to investigate the impact of temporal guidance and to explore the trade-offs between spatiotemporal and spatial-only visual guidance. Results show that the addition of temporal feedback makes visual guidance more effective. Our results reveal interesting design implications for future work in CVR and spatiotemporal guidance.

ACKNOWLEDGMENTS

We thank Ben Lafreniere and Frances Lai for initial discussions. We also thank our workshop and user study participants as well as anonymous reviewers for their invaluable feedback.

REFERENCES

[1] Activision. 2007. Call of Duty 4: Modern Warfare. [Playstation 3].
[2] Alcove VR. 2022. 360 3D MARSEILLE, FRANCE - Guided Tour - Virtual Travel 4K (Full series on Alcove for Quest). https://www.youtube.com/watch?v=gURIhqukTSQ. Copyright 2022 by Alcove VR. Reprinted with permission.
[3] Patrick Baudisch and Ruth Rosenholtz. 2003. Halo: a technique for visualizing off-screen objects. In *Proceedings of the SIGCHI conference on Human factors in computing systems*. 481–488.
[4] Carlo Bonferroni. 1936. Teoria statistica delle classi e calcolo delle probabilita. *Pubblicazioni del R Istituto Superiore di Scienze Economiche e Commericiali di Firenze* 8 (1936), 3–62.
[5] Felix Bork, Christian Schnelzer, Ulrich Eck, and Nassir Navab. 2018. Towards Efficient Visual Guidance in Limited Field-of-View Head-Mounted Displays. *IEEE Transactions on Visualization and Computer Graphics* 24, 11 (2018), 2983–2992. https://doi.org/10.1109/TVCG.2018.2868584
[6] Jessica Brillhart. 2016. In the Blink of a Mind — Attention. https://medium.com/the-language-of-vr/in-the-blink-of-a-mind-attention-1fdff60fa045 [Online; posted 05-Feb-2016].
[7] Gerd Bruder, Frank Steinicke, Phil Wieland, and Markus Lappe. 2012. Tuning Self-Motion Perception in Virtual Reality with Visual Illusions. *IEEE Transactions on Visualization and Computer Graphics* 18, 7 (July 2012), 1068–1078. https://doi.org/10.1109/TVCG.2011.274
[8] Stefano Burigat, Luca Chittaro, and Silvia Gabrielli. 2006. Visualizing Locations of Off-Screen Objects on Mobile Devices: A Comparative Evaluation of Three Approaches. In *Proceedings of the 8th Conference on Human-Computer Interaction with Mobile Devices and Services* (Helsinki, Finland) *(MobileHCI '06)*. Association for Computing Machinery, New York, NY, USA, 239–246. https://doi.org/10.1145/1152215.1152266
[9] Chong Cao, Zhaowei Shi, and Miao Yu. 2020. Automatic Generation of Diegetic Guidance in Cinematic Virtual Reality. In *2020 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*. 600–607. https://doi.org/10.1109/ISMAR50242.2020.00087
[10] Nina Doerr, Katrin Angerbauer, Melissa Reinelt, and Michael Sedlmair. 2023. Bees, Birds and Butterflies: Investigating the Influence of Distractors on Visual Attention Guidance Techniques. In *Extended Abstracts of the 2023 CHI Conference on Human Factors in Computing Systems* (Hamburg, Germany) *(CHI EA '23)*. Association for Computing Machinery, New York, NY, USA, Article 51, 7 pages. https://doi.org/10.1145/3544549.3585816
[11] Beat Games. 2018. Beat Saber. [Meta Quest].
[12] Beat Games. 2019. Beat Saber. [Meta Quest].
[13] Rockstar Games. 2008. Grand Theft Auto IV. [Playstation 3].
[14] Google. 2008. Google Maps. https://www.google.com/maps
[15] Google Spotlight Stories. 2016. 360 Google Spotlight Stories: HELP. https://www.youtube.com/watch?v=G-XZhKqQAHU.
[16] Steve Grogorick, Georgia Albuquerque, and Marcus A. Magnor. 2018. Comparing Unobtrusive Gaze Guiding Stimuli in Head-Mounted Displays. In *2018 IEEE International Conference on Image Processing, ICIP 2018, Athens, Greece, October 7-10, 2018*. IEEE, Athens, Greece, 2805–2809. https://doi.org/10.1109/ICIP.2018.8451784
[17] Steve Grogorick, Jan-Philipp Tauscher, Nikkel Heesen, Susana Castillo, and Marcus Magnor. 2020. Stereo Inverse Brightness Modulation for Guidance in Dynamic

Panorama Videos in Virtual Reality. In *Computer Graphics Forum*, Vol. 39. Wiley Online Library, 542–553.

[18] Uwe Gruenefeld, Abdallah El Ali, Susanne Boll, and Wilko Heuten. 2018. Beyond Halo and Wedge: Visualizing out-of-View Objects on Head-Mounted Virtual and Augmented Reality Devices *(MobileHCI '18)*. Association for Computing Machinery, New York, NY, USA, Article 40, 11 pages. https://doi.org/10.1145/3229434.3229438

[19] Uwe Gruenefeld, Dag Ennenga, Abdallah El Ali, Wilko Heuten, and Susanne Boll. 2017. EyeSee360: Designing a Visualization Technique for out-of-View Objects in Head-Mounted Augmented Reality. In *Proceedings of the 5th Symposium on Spatial User Interaction* (Brighton, United Kingdom) *(SUI '17)*. Association for Computing Machinery, New York, NY, USA, 109–118. https://doi.org/10.1145/3131277.3132175

[20] Uwe Gruenefeld, Ilja Koethe, Daniel Lange, Sebastian Weiß, and Wilko Heuten. 2019. Comparing Techniques for Visualizing Moving Out-of-View Objects in Head-mounted Virtual Reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 742–746. https://doi.org/10.1109/VR.2019.8797725

[21] Jan Gugenheimer, Dennis Wolf, Gabriel Haas, Sebastian Krebs, and Enrico Rukzio. 2016. SwiVRChair: A Motorized Swivel Chair to Nudge Users' Orientation for 360 Degree Storytelling in Virtual Reality. In *Proceedings of the 2016 CHI Conference on Human Factors in Computing Systems* (San Jose, California, USA) *(CHI '16)*. Association for Computing Machinery, New York, NY, USA, 1996–2000. https://doi.org/10.1145/2858036.2858040

[22] Sean Gustafson, Patrick Baudisch, Carl Gutwin, and Pourang Irani. 2008. Wedge: clutter-free visualization of off-screen locations. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems*. 787–796.

[23] Harmonix. 2005. Guitar Hero. [Playstation 2].

[24] Sandra G. Hart and Lowell E. Staveland. 1988. Development of NASA-TLX (Task Load Index): Results of Empirical and Theoretical Research. In *Human Mental Workload*, Peter A. Hancock and Najmedin Meshkati (Eds.). Advances in Psychology, Vol. 52. North-Holland, 139–183. https://doi.org/10.1016/S0166-4115(08)62386-9

[25] Indie Film School. 2016. Pokèmon 360 - CATCH 'EM ALL in VR! https://www.youtube.com/watch?v=pHUVS_GrIeM.

[26] Benjamin Jakobs. 2019. Die Entwickler des VR-Hits Beat Saber gehören jetzt zu Facebook. https://www.eurogamer.de/die-entwickler-des-vr-hits-beat-saber-gehoeren-jetzt-zu-facebook

[27] Oliver Beren Kaul and Michael Rohs. 2017. HapticHead: A Spherical Vibrotactile Grid around the Head for 3D Guidance in Virtual and Augmented Reality. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 3729–3740. https://doi.org/10.1145/3025453.3025684

[28] Joona Kinnunen. 2018. *Spatial guiding through haptic cues in omnidirectional video*. Master's thesis.

[29] Konami. 1998. Dance Dance Revolution. [Arcade].

[30] Daniel Lange, Tim Claudius Stratmann, Uwe Gruenefeld, and Susanne Boll. 2020. HiveFive: Immersion Preserving Attention Guidance in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems* (Honolulu, HI, USA) *(CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–13. https://doi.org/10.1145/3313831.3376803

[31] Jiannan Li, Jiahe Lyu, Mauricio Sousa, Ravin Balakrishnan, Anthony Tang, and Tovi Grossman. 2021. Route tapestries: Navigating 360 virtual tour videos using slit-scan visualizations. In *The 34th Annual ACM Symposium on User Interface Software and Technology*. 223–238.

[32] Klemen Lilija, Henning Pohl, and Kasper Hornbæk. 2020. Who put that there? temporal navigation of spatial recordings by direct manipulation. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems*. 1–11.

[33] Yen-Chen Lin, Yung-Ju Chang, Hou-Ning Hu, Hsien-Tzu Cheng, Chi-Wen Huang, and Min Sun. 2017. 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* (Denver, Colorado, USA) *(CHI '17)*. Association for Computing Machinery, New York, NY, USA, 2535–2545. https://doi.org/10.1145/3025453.3025757

[34] Yung-Ta Lin, Yi-Chi Liao, Shan-Yuan Teng, Yi-Ju Chung, Liwei Chan, and Bing-Yu Chen. 2017. 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* (Québec City, QC, Canada) *(UIST '17)*. Association for Computing Machinery, New York, NY, USA, 255–265. https://doi.org/10.1145/3126594.3126656

[35] Sean J. Liu, Maneesh Agrawala, Stephen DiVerdi, and Aaron Hertzmann. 2019. View-dependent video textures for 360° video. In *Proceedings of the 32nd Annual ACM Symposium on User Interface Software and Technology*. 249–262.

[36] Sandy Louchart and Ruth Aylett. 2003. Solving the Narrative Paradox in VEs – Lessons from RPGs. In *Intelligent Virtual Agents*, Thomas Rist, Ruth S. Aylett, Daniel Ballin, and Jeff Rickel (Eds.). Springer Berlin Heidelberg, Berlin, Heidelberg, 244–248.

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

[38] Carlos Marañes, Diego Gutierrez, and Ana Serrano. 2020. Exploring the impact of 360 movie cuts in users' attention. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. IEEE, 73–82.

[39] Carlos Marañes, Diego Gutierrez, and Ana Serrano. 2022. Towards assisting the decision-making process for content creators in cinematic virtual reality through the analysis of movie cuts and their influence on viewers' behavior. *International Transactions in Operational Research* (2022).

[40] Belen Masia, Javier Camon, Diego Gutierrez, and Ana Serrano. 2021. Influence of Directional Sound Cues on Users' Exploration Across 360° Movie Cuts. *IEEE Computer Graphics and Applications* 41, 4 (2021), 64–75.

[41] James McCrae, Michael Glueck, Tovi Grossman, Azam Khan, and Karan Singh. 2010. Exploring the Design Space of Multiscale 3D Orientation. In *Proceedings of the International Conference on Advanced Visual Interfaces* (Roma, Italy) *(AVI '10)*. Association for Computing Machinery, New York, NY, USA, 81–88. https://doi.org/10.1145/1842993.1843008

[42] James McCrae, Igor Mordatch, Michael Glueck, and Azam Khan. 2009. Multiscale 3D Navigation. In *Proceedings of the 2009 Symposium on Interactive 3D Graphics and Games* (Boston, Massachusetts) *(I3D '09)*. Association for Computing Machinery, New York, NY, USA, 7–14. https://doi.org/10.1145/1507149.1507151

[43] National Geographic. 2018. Elephant Encounter in 360 - Ep. 2 | The Okavango Experience. https://www.youtube.com/watch?v=HI7mTIxNotQ.

[44] Cuong Nguyen, Stephen DiVerdi, Aaron Hertzmann, and Feng Liu. 2017. Vremiere: In-headset virtual reality video editing. In *Proceedings of the 2017 CHI Conference on Human Factors in Computing Systems*. 5428–5438.

[45] Lasse T. Nielsen, Matias B. Møller, Sune D. Hartmeyer, Troels C. M. Ljung, Niels C. Nilsson, Rolf Nordahl, and Stefania Serafin. 2016. Missing the Point: An Exploration of How to Guide Users' Attention during Cinematic Virtual Reality. In *Proceedings of the 22nd ACM Conference on Virtual Reality Software and Technology* (Munich, Germany) *(VRST '16)*. Association for Computing Machinery, New York, NY, USA, 229–232. https://doi.org/10.1145/2993369.2993405

[46] Orbitian Media. 2022. One Day in London - VR/360° guided city tour (8K resolution). https://www.youtube.com/watch?v=v_4_gdbltyM. Copyright 2022 by Orbitian Media: https://orbitian.media. Reprinted with permission.

[47] Amy Pavel, Björn Hartmann, and Maneesh Agrawala. 2017. Shot orientation controls for interactive cinematography with 360 video. In *Proceedings of the 30th Annual ACM Symposium on User Interface Software and Technology*. 289–297.

[48] RenderHeads. 2022. *AVPro Video*. https://www.renderheads.com/content/docs/AVProVideo/articles/intro.html

[49] Sylvia Rothe, Felix Althammer, and Mohamed Khamis. 2018. 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* (Cairo, Egypt) *(MUM 2018)*. Association for Computing Machinery, New York, NY, USA, 115–119. https://doi.org/10.1145/3282894.3282903

[50] Sylvia Rothe, Daniel Buschek, and Heinrich Hußmann. 2019. Guidance in Cinematic Virtual Reality-Taxonomy, Research Status and Challenges. *Multimodal Technologies and Interaction* 3, 1 (2019). https://doi.org/10.3390/mti3010019

[51] Sylvia Rothe and Heinrich Hußmann. 2018. Guiding the viewer in cinematic virtual reality by diegetic cues. In *International Conference on Augmented Reality, Virtual Reality and Computer Graphics*. Springer, 101–117.

[52] School of Rock the Musical. 2015. SCHOOL OF ROCK: The Musical – "You're in the Band" (360 Video). https://www.youtube.com/watch?v=GFRPXRhBYOI.

[53] Ana Serrano, Vincent Sitzmann, Jaime Ruiz-Borau, Gordon Wetzstein, Diego Gutierrez, and Belen Masia. 2017. Movie Editing and Cognitive Event Segmentation in Virtual Reality Video. *ACM Trans. Graph.* 36, 4, Article 47 (jul 2017), 12 pages. https://doi.org/10.1145/3072959.3073668

[54] Alia Sheikh, Andy Brown, Zillah Watson, and Michael Evans. 2016. Directing attention in 360-degree video. (2016).

[55] Shakeeb Shirazi. 2018. *Timeline visualization of omnidirectional videos*. Master's thesis.

[56] Vincent Sitzmann, Ana Serrano, Amy Pavel, Maneesh Agrawala, Diego Gutierrez, Belen Masia, and Gordon Wetzstein. 2017. How do people explore virtual environments? *IEEE Transactions on Visualization and Computer Graphics* (2017).

[57] Marco Speicher, Christoph Rosenberg, Donald Degraen, Florian Daiber, and Antonio Krüger. 2019. Exploring Visual Guidance in 360-Degree Videos. In *Proceedings of the 2019 ACM International Conference on Interactive Experiences for TV and Online Video* (Salford (Manchester), United Kingdom) *(TVX '19)*. Association for Computing Machinery, New York, NY, USA, 1–12. https://doi.org/10.1145/3317697.3323350

[58] TL;DR Media. 2023. Simple Radar for Counter-Strike. https://readtldr.gg/simpleradar

[59] Jan Oliver Wallgrün, Mahda M. Bagher, Pejman Sajjadi, and Alexander Klippel. 2020. A Comparison of Visual Attention Guiding Approaches for 360° Image-Based VR Tours. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*. 83–91. https://doi.org/10.1109/VR46266.2020.00026

[60] Robert F Woolson. 2007. Wilcoxon signed-rank test. *Wiley encyclopedia of clinical trials* (2007), 1–3.

[61] Stefanie Zollmann, Tobias Langlotz, Raphael Grasset, Wei Hong Lo, Shohei Mori, and Holger Regenbrecht. 2021. Visualization Techniques in Augmented Reality: A Taxonomy, Methods and Patterns. *IEEE Transactions on Visualization and Computer Graphics* 27, 9 (2021), 3808–3825. https://doi.org/10.1109/TVCG.2020.2986247

A RADARVR IMPLEMENTATION DETAILS

RadarVR was prototyped in Unity with a Meta Quest 2 headset. The program plays 360° videos via RenderHeads's AVPro Video plug-in [48] and renders the RadarVR widget along the bottom edge of the viewer's display. For each 360° video, the Unity program takes as input a XML file with ROI labels. Each ROI label contains a bounding box (coordinates on the equirectangular video) and a start and end time.

We wrote custom shaders to animate and color the wedges and radar. As the viewer's head rotates, the radar visualization also rotates so that the front-facing direction coincides with the top of the radar. The distances of the wedges from the circle are synced with the video playback time, where the current playback time coincides with the circle circumference. Wedges appear 10 seconds before they reach the circle.

When an ROI is active (i.e., the corresponding wedge reaches the circle), the program checks whether the bounding box center is within the headset's field of view (FOV). Because the headset viewport is slightly wider than the display, RadarVR only marks an ROI as seen if the bounding box center is within the center 30% height and width of the viewport. When an ROI is marked as seen, the program gives feedback to the viewer by changing the color of the corresponding wedge to green and decreasing its opacity to $\alpha = 0.05$. If the viewer has oriented to the correct horizontal (yaw) direction but the ROI is above or below their view, the system displays a green triangle arrow above the radar to guide the viewer.

B USER STUDY EXPERIMENT VIDEOS

We selected five different 360° videos for our study. As shown in Table 2, the videos represent a range of genres and spatial and temporal distributions of ROIs. The video description and ROI labeling details are as follows:

France [2]. This virtual 360° tour shows various sites in Marseille, France. Each scene lasts about 20 seconds, where the narrator introduces the interesting tourist attractions and their history. We label regions that the narrator refers to (e.g., buildings, landscape) as ROIs. This medium-paced tour contains ROIs that are mostly spatially concentrated and are non-concurrent.

Elephants [43]. This National Geographic documentary tells a story about elephants in Botswana. In this slow-paced video, the video goes from close-up views of individual elephants, to migrating herds, and finally to general landscape and scenery. We label ROIs when elephants appear or perform some interesting action (e.g., charge at the camera), when the narrator refers to a specific regions in the scene, or if there are other relevant subjects or actions (e.g., camper starts cooking) in the story. Most ROIs are non-concurrent and are spread out in time and space.

Help [15]. This horror video tells a story of two main characters being chased by an alien, which grows larger and more menacing over time. During the chase, there were many moments where the alien was 180° opposite from the main characters, so the viewer has to turn frequently to see both the alien and the protagonists. We label salient parts of the story, such as important entrances (e.g., police officer appears) and events (e.g., alien breaks a window), as ROIs. This fast-paced story contains many short actions and events.

School of Rock [52]. This musical-style video takes place in a classroom. A teacher begins singing and incrementally adds students to the song. By the finale, everyone in the classroom participates and jams to the song. For ROIs, we label the main singer of each verse as well as interesting actions (e.g., teacher reveals a drum set). ROIs are usually non-concurrent but are frequent (i.e., verses in a medium-paced song). The finale involves many short phrases sung by different students across the room, so ROIs are short and spatially spread out towards the end.

Pokémon [25]. This 360° game-style video takes viewers to many different scenes and asks the viewer to find all the Pokémons in each scene. Some Pokémons appear for very short durations, while others appear for extended time periods. There are also Pokémons that appear high in the sky or near the feet of the viewer, which require viewers to look up or down. We label the main character and each Pokémon with an ROI for the duration of their appearance. There are many concurrent ROIs in this video, and their durations range widely.