Task-dependent Visual Behavior in Immersive Environments: A Comparative Study of Free Exploration, Memory and Visual Search

Sandra Malpica, Daniel Martin, Ana Serrano, Diego Gutierrez, and Belen Masia

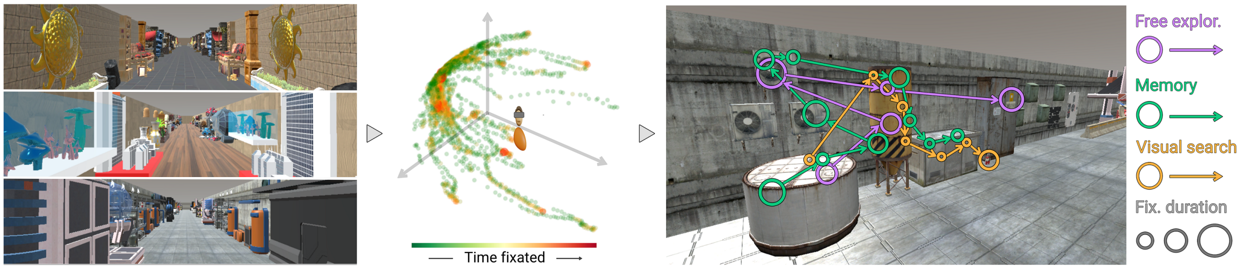

Fig. 1: We present a within-subjects systematic study of visual behavior in VR while conducting tasks with different cognitive loads: free exploration, memory, and visual search. We have designed three different scenes (*left*), and captured head and gaze information in 3D (*center*) from 37 participants performing the three different tasks in each of the scenes. Our analysis reveals significant differences on viewing behavior depending on the task: Free exploration (purple) yields longer, more spread fixations; memory (green) also exhibits long fixations, albeit closer in space; last, visual search (orange) elicits shorter fixations closer in space until the target element is found. The rightmost figure illustrates these different behaviors with data collected in our study.

Abstract—Visual behavior depends on both bottom-up mechanisms, where gaze is driven by the visual conspicuity of the stimuli, and top-down mechanisms, guiding attention towards relevant areas based on the task or goal of the viewer. While this is well-known, visual attention models often focus on bottom-up mechanisms. Existing works have analyzed the effect of high-level cognitive tasks like memory or visual search on visual behavior; however, they have often done so with different stimuli, methodology, metrics and participants, which makes drawing conclusions and comparisons between tasks particularly difficult. In this work we present a systematic study of how different cognitive tasks affect visual behavior in a novel within-subjects design scheme. Participants performed free exploration, memory and visual search tasks in three different scenes while their eye and head movements were being recorded. We found significant, consistent differences between tasks in the distributions of fixations, saccades and head movements. Our findings can provide insights for practitioners and content creators designing task-oriented immersive applications.

Index Terms—Virtual reality, attention, viewing behavior, task-dependent behavior, eye tracking.

1 Introduction

Creating appealing, engaging, and user-friendly experiences in VR requires understanding how users visually explore, observe, and interact with virtual environments. In particular, visual behavior (how we direct our gaze to our surroundings) is studied as a proxy to understand attention and other high-level cognitive processes [14], and often modeled following a bottom-up approach [25].

In order to capture visual behavior data, researchers often instruct participants to *freely explore* virtual environments [26, 39, 44, 52, 60]. However, day-to-day activities often require users to carry out different cognitive tasks, like memorizing visual content, or searching for a particular thing, which in turn affects visual behavior [53]. Kurz et al. [36] showed how most of the fixations cannot be explained only with external visual stimuli; instead, they tend to fall on task-related locations in the environment, limiting or even leading to the inhibition of bottom-up mechanisms [29, 56]. This has been largely supported by many other works, showing how gaze behavior in daily activities is strongly related to the evolution of the task itself [1, 20–22, 37, 38, 59].

Indeed, the differences between free exploration and task-dependent behaviors usually come from "just-in-time" mechanisms, from which information is picked depending on the task requirements at each moment [3]. Depending on the task, gaze can be directed towards novel areas, discarding already known information (as in a search task or while freely exploring), or go back to familiar locations, more efficiently avoiding unknown regions (as in a memorization task) [41].

However, we lack a systematic study of how different cognitive processes and tasks affect gaze in immersive environments. Existing works usually rely on different stimuli, methodologies, metrics and participants (for example, gaze behavior often exhibits considerable differences between users [55]), which makes drawing conclusions and comparisons between tasks difficult.

In this work, we conduct a systematic study on the differences in visual behavior between three different tasks: *free exploration*, where the participant does not have any particular goal, *memory*, which is linked to the processing and storage of information for a later use, and *visual search*, which is related to real-time information processing. These three tasks involve different cognitive processes, and are commonly used in applications like learning, videogames, or design [12, 40].

Our study features different scenes for a better generalization and is carefully designed to avoid repetitions while aiming for as similar as possible visual conditions between tasks. We also follow a within-subjects design and collect head and gaze information from all our participants. The analysis of our data (see Fig. 1) reveals that free exploration yields longer fixations, more spread across the space; memory shows a combination of short and long fixations, closer in space; and visual search elicits shorter fixations that follow a scanning strategy until the target is found. Our findings suggest that both bottom-up and top-down influences should be considered to help understand visual behavior. We found that, even though head movements vary significantly between tasks, eye eccentricity is not significantly affected by

- *Sandra Malpica, Daniel Martin, Ana Serrano, Diego Gutierrez and Belen Masia are with Universidad de Zaragoza - I3A. E-mail: smalpica | danims | anase | diegog | bmasia@unizar.es.*

such factors. This suggests that head data can be used as a proxy for gaze regardless of the cognitive task at hand or the visual content of the environment, a fact that had been previously reported only for free exploration of static 360º videos [58] and which may alleviate the dependence on real-time eye trackers in certain situations.

In summary, our contributions are:

- We provide a new dataset consisting of eye and head data recorded for 37 participants (we also provide the data of the 14 participants who carried out the pilot experiment) carrying out three different tasks in three different 3D scenes.
- We carry out a novel user study in a within-subjects design, gathering data that allows us to perform an analysis of task-dependent visual behavior.

Our results generalize across different scenes, despite their varying visual content. We believe this work represents a timely effort towards a better understanding of user visual behavior in VR, helping design task-oriented content. Our code and data can be found in https://graphics.unizar.es/projects/VR-TaskDependentGaze/.

2 Related Work

Despite a growing interest in visual behavior, only a limited number of studies have systematically investigated the differences across various tasks. In this section, we will first review works that specifically explore the three tasks we focus on, followed by a discussion of works that address multiple tasks, which are more closely related to our research.

Free Exploration Free exploration is a common task used to study visual behavior that has been widely explored in immersive environments during the last few years. Sitzmann et al. [58] introduced the first thorough study of visual behavior in immersive environments, finding biases and patterns in exploratory behaviors in static immersive environments under a free-viewing task. Since then, several studies have gathered large datasets of free-viewing behavior [9, 54]. Subsequent works have leveraged them to model visual attention, usually based on mechanisms such as visual saliency [6, 47] or scanpath prediction [2, 46]. More recently, these models have also incorporated auditory cues to account for multimodal attention [10, 66]. However, these models primarily rely on bottom-up processes and may not capture top-down factors that affect attention under different cognitive tasks. These limitations are particularly relevant for VR applications that involve underlying cognitive tasks. For instance, gaze patterns are crucial in cinematic VR [43, 57], where the user controls the visual focus in a 360-degree immersive environment. Therefore, it is critical to take into account top-down processes when developing attention models in VR, particularly when investigating more intricate tasks.

Memory This task involves important cognitive processes that allow us to encode, store, and retrieve information over time. It is an important task to explore in the context of immersive environments, as memory plays a vital role in our ability to navigate and interact with the world around us. Despite its importance, memory has not been studied extensively in immersive environments compared to free exploration. In 2D displays, Flanagan et al. [15] investigated the differences between the encoding phase and the recall phase by studying hand movements, while Hannula et al. [18] reviewed the relationship between gaze and memory tasks using fMRI. Kafkas et al. [31] focused on studying eye movements and their relation to recognition memory, discovering that pupil response and fixation patterns during memory encoding predict later memory strength. These studies, although not directly focused on eye behavior of memory tasks in immersive environments, provide valuable insights into how memory affects human behavior during different phases of the task.

Visual Search Investigating how humans actively seek out a target object among distractors provides valuable insights into the cognitive processes that drive visual perception. While visual search has been extensively studied in 2D traditional displays and in real life scenarios, its exploration in immersive environments remains limited. In 2D displays, several works have investigated human performance in different search tasks involving driving performance [62] or aviation performance [67]. Neider et al. [51] investigated how scene context guides eye movements during visual search. They studied differences in gaze patterns between constrained scenes (targets appear only in context-relevant locations) and unconstrained scenes (targets appear anywhere in the scene), and showed that there are important differences in eye metrics such as dwell time and number of fixations. In immersive environments, Enders et al. [13] studied gaze behavior during navigation and visual search and found that fixation rates onto targets decrease when participants have to perform complex tasks. Although these works provide relevant and interesting insights, they often focus on very specific visual search tasks and scenarios. To gain a comprehensive understanding of how humans perform visual search in virtual environments, it is essential to identify common gaze patterns across more general and diverse immersive scenarios.

Multiple tasks A few works have studied differences in gaze patterns when performing different tasks. In the context of 2D displays, Bryan et al. [8] compared observation, recall and query answer (search) tasks on chart visualization, finding that the search task led to gaze patterns that were significantly different from the other two. Regarding immersive environments, Li et al. [40] studied how memory affects visual search in 2D and 3D scenarios. They found that memory helps to allocate search to a restricted part of the environment. Kit et al. [33] studied the relation between visual search and scene memory in a visual search task over three days. They found that, similar to two-dimensional contexts, participants quickly learned the location of targets in the environment over time, and used spatial memory to guide search. Different cognitive tasks also have an impact on visual attention [61], as well as carrying out more than one task at the same time like walking and memorizing [50].

Closer to our work, Hadnett-Hunter et al. [17] investigate the effect of the task on visual attention in interactive virtual environments. In particular, they compare free viewing, search and navigation tasks by carrying out a user study using traditional displays and a chin rest. It is important to note that this set up restricted the head movements of the participants, which are an important part of visual behavior, as discussed by Hadnett-Hunter et al. The interaction with the virtual environment was supported by the use of a keyboard and a mouse. In contrast, we study the effect of cognitive tasks on visual behavior in an immersive environment, using a VR headset which directly transfers the participants' physical movement into the virtual world, allowing for natural head movements and a more natural interaction with the environment. Additionally, the different tasks of the experiment carried out by Hadnett-Hunter et al. have different durations, and different starting or (in the case of navigation) ending points. This means that the participants are not subject to the same visual content while carrying out the different tasks. In comparison, our experiment is designed in such a way that it ensures a uniform duration for the different tasks, and that the overall visual information that participants perceive while carrying out different tasks is as uniform as possible. Regarding the findings, we both find significant differences on gaze depending on the task.

Hu et al. [24] present a learning-based method (EHTask) to recognize user tasks in VR. Although they are mainly focused on this method, they are also the first to collect a dataset of eye and head movements for different cognitive tasks in an immersive environment. They compare free viewing, visual search, saliency and tracking using monoscopic 360-degrees videos. Hu et al. discuss how the use of 2D, non-interactive stimuli is a limitation of their work compared to more natural virtual environments. In contrast, our experiment allowed for more complex and natural visual cues like binocular disparity and motion parallax. Additionally, Hu et al. present the same video four times to each user (once for each task) without accounting for possible repetition effects, which are known to affect visual behavior [5]. On the other hand, our experiment is designed to avoid exact repetitions while trying to achieve uniform visual conditions between tasks. Regarding our findings, we both find a significant effect of the different tasks on both eye and head movements, which confirms that head movement is an important part of visual behavior.

While these works provide valuable insights into human behav-

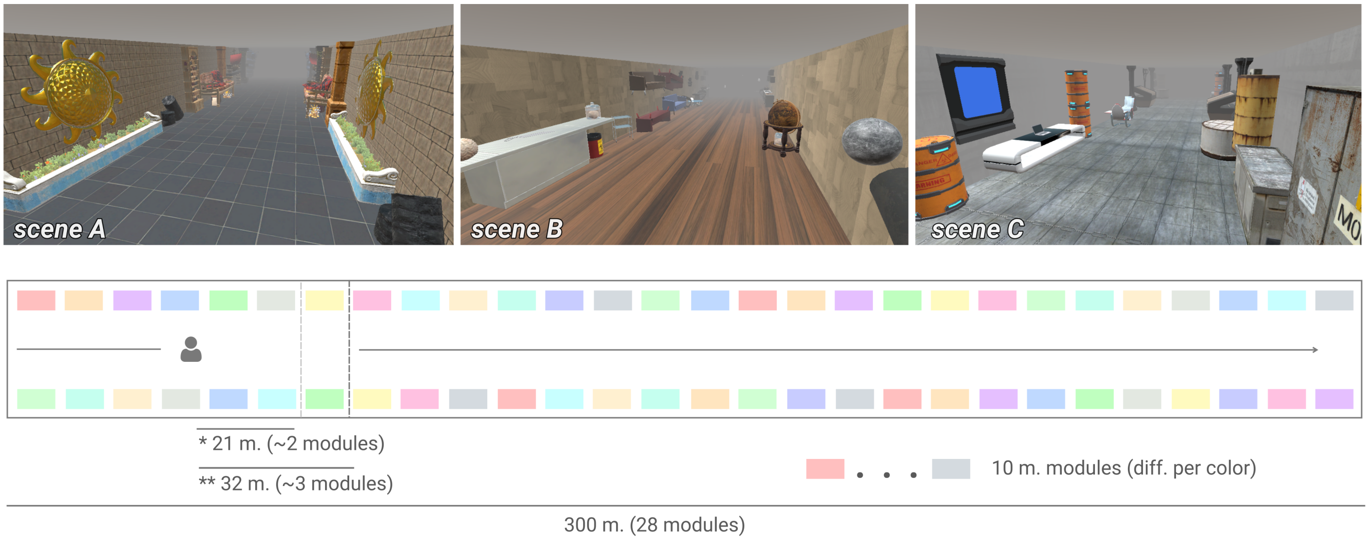

Fig. 2: *Top*: Sample views of each of the three scenes used in our main experiment. *Bottom*: Zenithal schema of the corridors used as scenes in our main experiment. The corridors are 300-meters long, and built by pseudo-randomly arranging a pre-designed set of fourteen 10-meter modules (repetition in contiguous modules is avoided). The pre-designed set of modules is different for each of the three scenes. The participant moves forward along the center of the corridor at a constant speed of 2.5 m/s riding on a virtual wagon while seated in a physical office chair. (*) A dynamic fog starts appearing at 21 meters (two modules) away from the user, (**) completely covering vision at 32 meters (three modules) away from the user.

ior during different tasks, they often investigate these tasks in a non-systematic way, using different stimuli, participants and metrics. This can make it challenging to draw meaningful conclusions or comparisons between tasks and studies. In contrast, our study employs a systematic approach to investigate how different cognitive tasks impact visual behavior in immersive environments. We use a novel within-subjects design scheme, carefully designed to provide as similar conditions as possible between tasks while avoiding exact repetitions in order to ensure a comprehensive and robust investigation of visual behavior across different tasks.

Table 1: Variables analyzed in our pilot and main user studies.

Variable (unit)	Comments
Number of fixations (per sec.)	Averaged from the 60s trial
Duration of fixations (sec.)	Mean duration of fixations
Dwell time (sec.)	Sum of fixations' durations
Number of saccades (per sec.)	Averaged from the 60s trial
Duration of saccades (sec.)	Mean duration of saccades
Amplitude of saccades (deg)	Mean amplitude of saccades
Eye eccentricity (deg)	Mean gaze eccentricity
Head orientation entropy	Mean Shannon entropy of head

3 Experimental Design

We designed a study to analyze *how the task being carried out by a user affects their viewing behavior in immersive environments*. Specifically, we examined three different tasks: a *free exploration* task, in which participants advanced through the scene without any specific instructions; a *memory* task, in which they had to remember the scene in order to answer questions about it afterwards, and a *visual search* task, in which participants were instructed to look for a particular object.

Our study was carefully designed to feature different visual scenes and use a within-subjects design while avoiding repetitions. Part of our design decisions were tested by means of a pilot experiment, prior to the main experiment. Therefore, the present section describes the common aspects of the experimental design. Meanwhile, Sec. 4 describes the specifics of the pilot and main experiments, and presents and discusses their results.

3.1 Study Variables

Given the aim of our study, our main factor (independent variable) is the *task* being conducted by the participant, with three levels: *free exploration*, *memory*, and *visual search*. Additionally, and in order to improve the generality of our findings, our main experiment measured viewing behavior in three different scenes that exhibit a variety of objects and appearances, yielding a second independent variable with three levels, one per scene, in the main experiment.

Since we (i) did not want to tie a task to a specific scene, but rather seek a full combinatorial design, and (ii) chose a within-subjects design in which the same participants perform the various tasks, there was a risk that being subjected to the same scene several times affected visual behavior. For example, fixation duration tends to decrease when repeated stimuli are presented [5, 30, 32]. Repetitions can also decrease the number of fixations regardless of the demographic differences of participants [16]. To avoid repetitions of the visual content while aiming at having similar visual conditions between tasks we employed *variations* of the scenes, designed so that they have the same visual information, albeit in a different spatial arrangement. We describe how these variations are generated in Sec. 3.2. To make sure our variations in spatial arrangement were not affecting viewing behavior, we conduct a pilot experiment testing this (Sec. 4.1), prior to the main experiment (Secs. 4.2 and 4.3).

As for our dependent variables, we establish the following eight variables (see Table 1), which are widely-used measures representative of viewing behavior in immersive environments [13, 58] and can be computed from the recorded head and gaze data: number and duration of fixations, number and duration of saccades, amplitude of saccades, dwell time, eye eccentricity and head orientation entropy. Classification of eye movements into fixations and saccades was done using PyGaze [11] (with a maximal inter-sample distance of one visual degree in the detection of fixations, a velocity threshold of 30 deg/s in the detection of saccades, and a minimal duration of 50 ms for both). Fixation duration was computed as the mean duration of fixations in a single trial, while dwell time was computed as the summation of the duration of all fixations in a trial. Eye eccentricity is given by the distance, in visual degrees, between the gaze point and the viewport center [58]. Head orientation entropy was computed as the Shannon entropy of the orientation of the head in the virtual world. We also gathered information on sickness and presence, through questionnaires. While our analysis focuses on viewing behavior, we did want to ensure our participants did not exhibit severe sickness symptoms, which could affect gaze behavior. None of our participants reported moderate or severe sickness symptoms. There were no differences between the tested conditions in our studies (pilot or main experiment) regarding sickness.

3.2 Stimuli

We used corridors populated with objects (see Fig. 2, top row, and Fig. 3) as our experimental scenario. The virtual camera advanced along the corridor with constant velocity and rectilinear motion (the participant experienced it as traversing the corridor while on a platform with wheels). This choice of scenario allowed us to: (i) provide rich variety in the objects present in the scene, necessary for our tasks, (ii) have a dynamic setup, closer to a natural 3D scenario and (iii) constrain the participants' translation in the virtual world (instead of letting them freely move around), so that they were all exposed to the same scenery.

Each of our *scenes* was therefore a corridor, with elements and objects of a certain style. Corridors were 14 meters wide and 300 meters long, and contained objects arranged close to both walls. We used 3D models obtained from the Unity Asset Store, additional open-source repositories, or modeled in Unity by the experimenters.

For each scene, we built different 10-meter-wide modules; 28 modules were fit in each wall of the corridor, for a total of 56 modules per scene. Each unique module was included several times within its scene, ensuring no two identical modules were contiguous. In order to implement the *variations* of each scene introduced in Sec. 3.1 we randomly shifted the modules to produce different spatial arrangements while maintaining the constraint of no contiguous repetition.

To traverse the corridor in a controlled manner (as explained at the beginning of this subsection), participants were placed in a virtual wheeled platform that moved steadily at 2.5m/s. We sought to analyze temporal windows of 60 seconds, which is on par with previous literature [9,65]. We gathered two minutes of data to anticipate any potential hurdle during data capture, e.g., due to participant issues or equipment malfunction. To avoid participants focusing on areas that were too far away, a gray fog started appearing at 21m from the participant's position, and obstructed the view completely at 32m (see Fig. 2, bottom row). This is representative of (equal or larger than) a viewing range in depth common in many scenarios, and in our case it meant that participants could always clearly see at least two modules in front of them in any direction. Fig. 1 (left) shows a sample view of the scenes without this fog.

3.3 Hardware

The stimuli were presented on an HTC Vive Pro Eye head-mounted display (HMD) with a horizontal field of view (FoV) of 110 visual degrees and a vertical FoV of 110 visual degrees, a resolution of 1440 x 1600 pixels per eye, and a frame rate of 90 frames per second. The HTC sensors to track participants' head position and the integrated eye-trackers to record gaze information worked at the same frequency. Participants used the HTC Vive Controller to provide trial responses when needed (Sec. 3.4 includes details on the trials and tasks). We logged head and gaze information during the whole experiment.

3.4 Procedure: General Aspects

We describe here the aspects of the procedure that are common to *both* our pilot study and our main experiment, whereas the particularities of each are described in Sec. 4.1 and Sec. 4.2, respectively.

Participants were seated on a swivel chair, which allowed them to turn and look around if they wished, but they could not translate. The HMD was used in tethered mode, and the cables were positioned so that they didn't interfere with their actions. First, the experimenter explained to them how to properly adjust the HMD and gave instructions on the experiment, including how the controllers worked. Once the HMD was properly adjusted, calibration of the eye tracker was performed, and then the experiment started. At the end of the experiment, participants filled in demographics, sickness and presence questionnaires (please refer to the supplementary material for these questionnaires), and an informal debriefing session was conducted.

During the experiment, in both the pilot and the main experiment participants, one *trial* was the complete traversal of a corridor by the participant, performing one of the three selected tasks: *free exploration*, *memory*, or a *visual search* task. In the free exploration task the participants only had to watch the corridor as if walking through, with no particular purpose other than exploring. In the visual search task, participants were instructed to press the trigger of the controller every time they saw a specific object (candles in *scene A*, brains in *scene B* and fire extinguishers in *scene C*). In the memory task, they were instructed to memorize what they saw to answer a question at the end of the trial (e.g., "Name five specific objects you remember seeing in the scene" or "Name five characteristics of an object you remember seeing in the scene").

The whole experimental procedure for our study was approved by the Research Ethics Committee of the Autonomous Community of Aragon, Spain (CEICA). All participants voluntarily took part in the study and provided written consent for participation, knowing they could stop the experiment at any moment if they wished to do so.

4 Effect of Cognitive Task on Visual Behavior

Our study consists of a pilot experiment and a main experiment. The purpose of the pilot experiment was to test the effect of different spatial variations (introduced in Sec. 3.1) of the same visual content on visual behavior, in order to avoid repeating exactly the same scene when the users were carrying out different tasks. We describe this pilot and its findings in Sec. 4.1. Once that is asserted, we move on to describing the main experiment and its findings (Secs. 4.2 and 4.3). Since Sec. 3 describes common aspects pertaining to both the pilot and main experiments, the present section focuses on the particularities of each experiment, as well as describing the results and discussing the findings.

4.1 Pilot Experiment: Effect of Spatial Arrangement on Visual Behavior

Stimuli We built a scene for the pilot, following the same structure described in Sec. 3.2 (same size and layout), but different from the three scenes used in the main experiment. We generated seven unique 10-meter-wide modules for this scene, and included each module eight times within each corridor, ensuring no identical modules were contiguous. Three variations of this scene were generated by randomly shuffling the modules, while maintaining the constraint of no contiguous repetition. Two views from different variations of the scene can be seen in Fig. 3 (left).

Participants The pilot experiment was carried out by fourteen participants (eight identified themselves as female, six as male, and none as non-binary, other, or preferred not to say; average age 24.92 years old, STD = 4.87). They all reported normal or corrected-to-normal vision, and were naïve about the final purpose of the experiment.

Procedure Each participant carried out one task (*free exploration*, *memory*, or *visual search* as described in Sec. 3.4) in the three different variations of our scene, thus doing three trials (3 *variations* × 1 *scene* × 1 *task*). Task assignment to a participant was pseudo-randomly done so that it was balanced across participants, and within each participant the three variations of the scene were shown in random order. The apparatus used was as described in Sec. 3.3.

Statistical analysis The pilot follows a between-within subjects design. We establish the significance level at $p = 0.05$ and employ a generalized linear mixed model (GLMM), since it provides a robust and flexible approach for analyzing non-normal data when random effects are present [7]. There are three independent variables: *task* (between-subjects, with 3 levels: *free exploration* vs. *memory* vs. *visual search*); *variation* (within-subjects, with 3 levels: the three different spatial arrangements of the scene) and *order* (within-subjects, with 3 levels: the order of presentation–1st, 2nd or 3rd). The statistical analysis was carried out using Matlab [48].

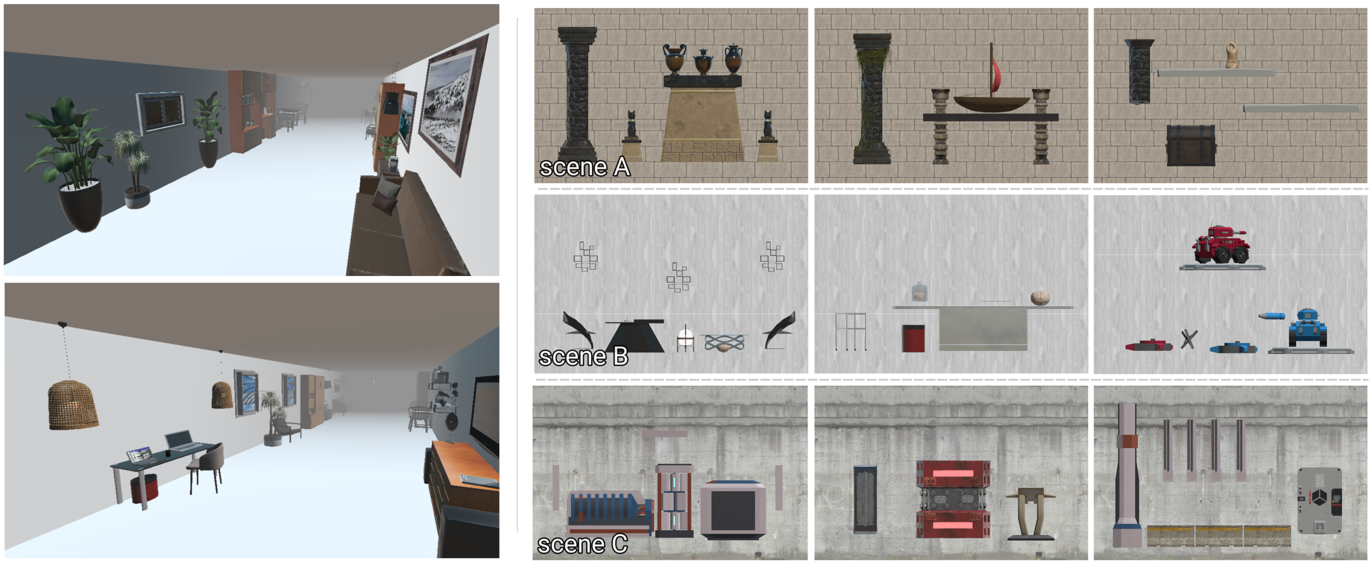

Fig. 3: *Left*: Two views of different variations of the scene used in our pilot experiment. *Right*: We show three sample modules belonging to each of the three different scenes (one per row) that we designed for our main experiment. Each module was 10-meters long. Note that these modules were also designed to have the regions of interest in different locations, thus potentially eliciting different viewing patterns. Representative views of those scenes can be found in Fig. 2.

Table 2: Results from the statistical analysis of our pilot experiment. We have found that neither the *order*, the *variation* of the scene, nor their interaction had a significant effect on any of our analyzed variables (all p-values are notably above the significance threshold of $\alpha = 0.05$). This suggests that using different variations of the same scene does not effectively affect viewing behavior, and thus may be used to diminish habituation. Further details can be found in Sec. 4.1.

	Order		**Variation**		**Order:Variation**	
Variable	**t-stat**	**p-value**	**t-stat**	**p-value**	**t-stat**	**p-value**
Number of fixations	-0.248	0.806	-0.508	0.616	0.613	0.546
Duration of fixations	0.108	0.913	1.084	0.278	-0.636	0.524
Dwell time	0.093	0.925	0.066	0.947	-0.073	0.941
Number of saccades	0.001	0.998	-1.077	0.293	0.315	0.755
Duration of saccades	0.163	0.870	1.057	0.290	-0.205	0.837
Amplitude of saccades	1.540	0.123	-0.064	0.948	-0.367	0.713
Eye eccentricity	0.068	0.945	0.542	0.590	-0.004	0.996
Head entropy	-0.408	0.685	-0.102	0.918	0.180	0.857

Research hypotheses Our main hypothesis is that changing the spatial arrangement of objects within the scene will not have a significant effect on visual behavior while avoiding the potential effects caused by repetitions. Thus, we hypothesize that the scene *variation* will not significantly affect the eight variables representing visual behavior (described in Sec. 3.1). Since the task is a between-subjects factor in this experiment, and it is well-known that visual behavior can vary significantly between users, we do not look into the effect of *task* here.

Results We analyzed the first 60s of each trial for every participant. We found that neither the *order* nor the *variation* had a significant effect on any of the analyzed dependent variables descriptive of visual behavior (see Table 2). We thus decided to use different variations of the scenes in our main experiment in order to avoid repetitions. We assume that these variations can be used as a way to avoid potential effects of repetition or habituation in our main experiment while keeping a similar visual content, since participants will see the same objects with different spatial arrangements. Studying how repeated stimuli affect different tasks can lead to further insights on how different cognitive processes like novelty or habituation affect visual behavior, but is out of the scope of this paper. Please refer to the supplementary material for the full results of the analysis.

4.2 Main Experiment: Description

Stimuli The main experiment featured three different scenes (*scene A*, *scene B* and *scene C*, showing ancient, contemporary or futuristic objects), following the size and layout described in Sec. 3.2. We built fourteen unique 10-meter-wide modules for each scene, and included each module four times within each corridor, ensuring no identical modules were contiguous. Three variations of each scene were generated by randomly shuffling the modules, while maintaining the constraint of no contiguous repetition. A variation of each scene is shown in Fig. 2 (top row), while various sample modules of each scene can be seen in Fig. 3 (right).

Participants Our main experiment was carried out by 37 participants (18 identified themselves as female, 18 as male, none as non-binary, and one preferred not to say; average age 33.56 years old, STD = 15.80). None of them had participated in the pilot experiment. They all reported normal or corrected-to-normal vision, and were naïve about the final purpose of the experiment.

Procedure As explained, we chose a within-subjects design in our main experiment, motivated by the fact that gaze behavior often exhibits considerable differences between users while keeping more stable gaze patterns within users [55]. Therefore, each participant performed all three tasks (described in Sec. 3.4) in each of the three different scenes, leading to nine trials per participant (3 *tasks* × 3 *scenes*). All participants carried out each task randomly in a different variation (different spatial arrangement of the modules) of the same scene in order to avoid exact repetitions of the visual content. The order of the nine trials was randomized. These nine test trials, featuring unique corridors, were complemented with a re-run trial. For each participant, this re-run trial was equal to the first visual search trial they had experimented, repeated again at the end of the nine test trials, as an additional measure of intra-user congruency. The experiment had a total duration of twenty minutes.

Statistical analysis The main experiment follows a full-combinatorial within-subjects design with two factors: *task* (3 levels) and *scene* (3 levels). As in the analysis from our pilot experiment (Sec. 4.1), our data is not normally distributed and includes random effects (e.g., participant). We thus conduct a 3 task (*free exploration* vs. *memory* vs. *visual search*) × 3 scenes (*scene A* vs. *scene B* vs. *scene C*) generalized linear mixed model (GLMM) to analyze the effects of such factors, and their interactions, on the dependent variables. Our

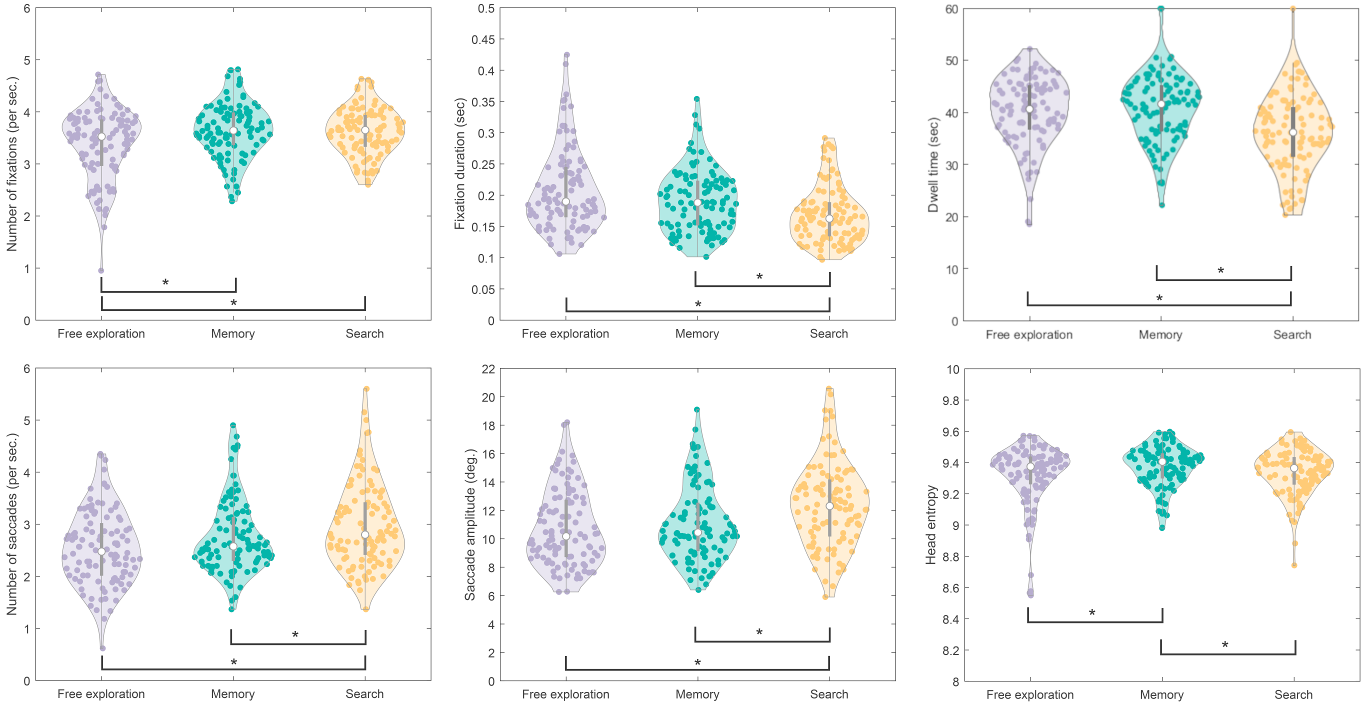

Fig. 4: Results of our main experiment for the three tasks: free exploration (purple), memory (green) and visual search (yellow). Each set of three violin plots correspond to a dependent variable analyzed; in reading order: number of fixations (per second), fixation duration, dwell time per trial, number of saccades (per second), saccade amplitude and head orientation entropy. Statistically significant differences according to the post-hoc tests are marked with an asterisk.

Table 3: Median value and 95% confidence interval for each of the dependent variables in our main experiment, per task. Variables that are significantly affected by task are marked in bold. Post-hoc significant differences between tasks are marked with symbols (○ and ×), such that, for each row, tasks with the same symbol are not significantly different between them, and tasks with different symbols are.

Variable	**Free exploration**		**Memory**		**Visual search**	
Number of fixations (per sec.)	3.461 ×	[3.392, 3.529]	3.569 ○	[3.476, 3.662]	3.523 ○	[3.454, 3.593]
Duration of fixations (sec.)	0.192 ○	[0.189, 0.195]	0.177 ○	[0.175, 0.179]	0.155 ×	[0.154, 0.156]
Dwell time (per trial)	40.658 ○	[40.098, 41.218]	41.629 ○	[41.117, 42.141]	36.189 ×	[35.610, 36.768]
Number of saccades (per sec.)	2.005 ○	[1.976, 2.034]	2.022 ○	[1.997, 2.048]	2.081 ×	[2.051, 2.110]
Duration of saccades (sec.)	0.084 ○	[0.083, 0.085]	0.083 ○	[0.083, 0.084]	0.089 ○	[0.089, 0.900]
Amplitude of saccades (deg.)	9.090 ○	[8.990, 9.190]	10.156 ○	[10.061, 10.251]	10.941 ×	[10.826, 11.056]
Eye eccentricity (visual degrees [58])	6.202 ○	[6.139, 6.265]	6.389 ○	[6.326, 6.453]	6.216 ○	[6.151, 6.282]
Head orientation entropy	9.374 ○	[9.364, 9.385]	9.405 ×	[9.395, 9.415]	9.364 ○	[9.358, 9.370]

dependent variables are the measures we choose as descriptive of visual behavior (see Sec. 3.1 for more details on their computation): (i) number of fixations per trial, (ii) number of saccades per trial, (iii) dwell time per trial, (iv) duration of fixations, (v) duration of saccades, (vi) amplitude of saccades, (vii) eye eccentricity, and (viii) head orientation entropy (for all variables except (i)-(iii), the analysis is done on the average value per trial[1]). Note that, for each trial, we analyze the first 60 seconds for every participant. For each of these analyses, we additionally conduct post-hoc pairwise Bonferroni-corrected comparisons. All statistical analyses were carried out using Matlab. Effect sizes are calculated using partial omega squared [64] which is suitable for a non-parametric analysis like the GLMM. We consider small effect sizes $\Omega_p^2 < 0.01$, medium effect sizes $0.01 \leq \Omega_p^2 \leq 0.06$ and big effect sizes $\Omega_p^2 > 0.06$, and establish significance level at $p = 0.05$. Post-hoc analyses were carried out via marginal means with the emmeans Matlab package.

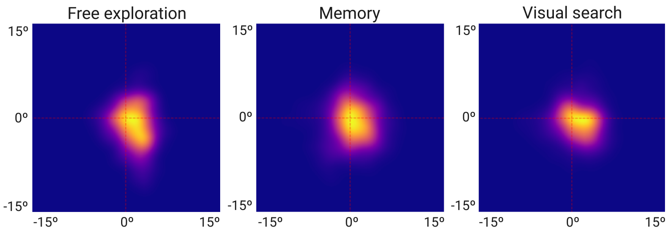

Fig. 5: Aggregated eye eccentricity for each of our three tasks. For each trial, task and participant, we split the data into one-second windows, and compute the average eccentricity per window. Heatmaps show the aggregation of these average eccentricities, separated by task. While slight differences between tasks are observable, our analysis reveals no statistically significant effect of the task on eye eccentricity.

4.3 Main Experiment: Results and Discussion

In the following, we discuss the main significant results indicating whether and how the task affects visual behavior. The full results of the statistical analysis can be found in the supplementary material. Regarding the performance in the task, mean accuracy in visual search was 88.98% and all 37 participants were able to recall five objects or five characteristics as asked in every memory trial.

[1] This is also the case in the pilot experiment.

Effects on gaze behavior We describe here effects of the task on measures related to ocular movements (fixations and saccades); these measures include: dwell time, duration and number of fixations; and duration, amplitude and number of saccades. Fig. 4 plots the main results and significant differences between tasks, while Table 3 shows mean values and 95% confidence intervals (CIs) per task for all dependent variables.

We find significant main effects of *task* on all measures related to fixations: number and duration of fixations, as well as dwell time. In particular, *free exploration* has significantly less fixations than *memory* and *visual search* ($t = 6.393$, $p < 0.001$, $\Omega_p^2 = 0.016$). The duration of fixations is also affected by task ($t = 5.890$, $p < 0.001$, $\Omega_p^2 < 0.024$), with *visual search* leading to significantly shorter fixations than the other tasks. If we consider dwell time, there is also a significant, large effect of task ($t = 4.115$, $p < 0.001$, $\Omega_p^2 < 0.078$) with *visual search* having a lower total dwell time. Regarding saccades, it is interesting to see that the duration of saccades is not affected by the task. The number of saccades, however, does exhibit a significant medium-sized effect of task ($t = 14.00$, $p < 0.001$, $\Omega_p^2 < 0.049$), and the amplitude of saccades a significant effect of task ($t = 6.514$, $p < 0.001$, $\Omega_p^2 < 0.065$). We find significantly more saccades, and of larger amplitude, in the *visual search* condition, consistent with the fixation-related metrics described above, which already suggested a more frequent shift of visual attention.

Our results suggest that participants *shift their focus more often* in the visual search task—and to a lesser extent in the memory task too, trying to register as much of their surroundings as possible in a fast way. The fact that the participants are moving forward at a steady velocity during the experiment means their time is limited, and fast shifts of attention may be a strategy to optimize their task performance. Additionally, the shorter fixations in the visual search task may be caused by participants losing interest in an object as soon as they perceive it is not their search target, following a serial deployment of attention [63]. Conversely, fixations are longer in the free exploration task, where participants can process more information from each object that captures their attention, following the so-called focal exploration mode [49].

Effects on eye eccentricity We found no significant effect of task ($t = 0.2132$, $p = 0.831$), scene ($t = -0.905$, $p = 0.366$), or any of their interactions on eye eccentricity. Fig. 5 shows aggregate eye eccentricities per task, showing that indeed values are similar across tasks; we also observe a slight bias towards the right, which may be owing to predominant participant right-handedness. This independence of eccentricity with the task being conducted can be beneficial for the use of head movements as a proxy for gaze behavior when eye tracking is not available. Previous works have assessed the feasibility of this in free exploration contexts [23, 58], e.g., by using head position while fixating convolved with a Gaussian kernel whose size was based on observed eye eccentricity as the input to compute a *head saliency map* of the scene.

Effects on head orientation We also studied the entropy of the head orientation and found a significant effect of task ($t = -2.077$, $p = 0.0385$, $\Omega_p^2 = 0.039$) and scene ($t = -2.199$, $p = 0.0286$, $\Omega_p^2 = 0.025$). In general, the entropy of the head orientation is higher for the *memory* task than for the *free exploration* and *visual search* tasks, with no significant difference between the last two (see Table 3). A possible explanation for this is that participants move their head more (trying to cover a higher percentage of the environment), and in a less uniform manner, when they have to remember the scene which leads to higher entropies.

Influence of visual content Presenting our participants with three different scenes serves a double purpose: first, it decreases learning or habituation effects, and second, it provides generality to our findings beyond a single specific layout. We analyze the effect of visual content (scene factor) on our dependent variables, and observe no significant effect of scene except in two variables: fixation duration ($t = 2.826$, $p = 0.005$, $\Omega_p^2 < 0.001$) and head orientation entropy ($t = -2.200$, $p = 0.029$, $\Omega_p^2 = 0.017$). In the case of fixation duration, the effect size is very small, while in the case of head entropy it is small to medium-sized effect, and we observe a higher entropy in the case of *scene C*, which we attribute to its higher visual complexity. Nevertheless, it is important to note that the effect of different tasks on visual behavior is consistent regardless of the differences between scenes, which suggests our findings can be of a more general nature.

5 Conclusion

In this work we have presented a novel study that investigates how performing different cognitive tasks (*free exploration*, *memory*, and *visual search*) influences visual behavior in 3D immersive environments. While previous work has addressed visual behavior in everyday activities [34], we have resorted to VR since it offers a much more controlled environment, reduced confounding factors, and easier stimuli manipulation, which are key components in experiments like ours. Moreover, VR technology has gained significant traction in the last few years, albeit with viewing conditions that still differ in some aspects from those in the real world (e.g., regarding aspects of visual acuity [27], or the accommodation-vergence conflict [35]). Our study is unique in that we have measured task-dependent differences within the same subjects, who performed all three tasks in three different 3D immersive scenarios. This approach has enabled us to conduct extensive analyses on how distinct inherent features of visual behavior are impacted by the specific task being performed [58].

We have found visual search to be significantly different from the other two tasks, with participants rapidly scanning the scene, moving from one element to the next trying to find some target object. This translates into devoting significantly less time in fixating, while performing more, ampler saccades, to faster reach such target. Note that this behavior may be related to the experimental set up itself. Since the participants are moving at a constant, fixed velocity, they are forced to scan their surroundings within a certain amount of time, thus shortening fixation duration. For example, Hu et al. [24] find that fixation duration is significantly shorter in the free viewing condition than in the visual search condition under 360-degrees video visualization, while Hadnett-Hunter et al. [17] find no significant difference in fixation duration between free viewing and visual search using traditional displays and a chin rest. It is thus important to consider how the overall set up or final application can modulate the effect of the different tasks. We chose a smooth, limited movement in order to (i) reduce potential sickness in our application [42] and (ii) to ensure that every participant followed the same trajectory and was exposed to the same visual content. However, it would be interesting to study how free movement in an immersive, 3D environment further affects visual behavior. Our results also show that participants performing free exploration incur in fewer, yet longer fixations. This suggests that in the absence of a clear task, participants tend to focus on elements that capture their attention. In our study, we observed that eye eccentricity did not differ significantly across tasks. This suggests that head movement can serve as a reliable proxy for gaze to reduce the dependence on real-time eye trackers in immersive applications, as pointed out by Sitzmann et al. in the context of free exploration of static 360° panoramas, regardless of the cognitive task being performed. Our analyses revealed that the average eye eccentricity across tasks was 6.34°, which falls within the parafoveal region where humans can still see clearly [4].

Regardless of the task, there are many other factors that can also affect visual behavior and could be explored in future works. For instance, visual behavior in a given context can change over the course of time or exhibit systematic tendencies and biases. The temporal dynamic characterization of visual behavior distinguishes two different viewing strategies: ambient and focal [39]. The focal strategy is characterized by shorter saccades and is used to focus on specific and adjacent locations, while the ambient strategy helps obtain a global understanding of the environment with longer saccades. The ambient strategy is usually predominant during the first few seconds of a task, while the focal strategy is used more often and it is also more driven by bottom-up saliency [49]. A complete temporal analysis of our data could yield deeper information about the differences or similarities of the studied task. For example, perhaps the ambient strategy is used more often in our visual search task, which could also be related to the larger amplitude of saccades we found. It is also possible that the different tasks differ more in one of the strategies, or that the relative time of each strategy varies depending on the task the participant is carrying out.

The potential applications of our findings are numerous, from enabling task recognition (as suggested by Hu et al. [24]) or saliency prediction (as observed by Hadnett-Hunter et al. [17]) in interactive environments, to informing the design of more effective and enjoyable immersive experiences [17]. Task-specific visual behavior features can be used to identify the task users are performing, e.g. in video games or other interactive environments, in order to improve user experience and performance; this can be done, for instance, by helping the user conduct the specific task if they are taking too long. Further, existing approaches that focus on modeling and predicting visual behavior in immersive environments [6, 28, 46] are primarily trained on datasets containing viewing data from free exploration tasks, making them less effective in modeling behaviors related to other tasks. Looking ahead, we believe that different gaze prediction models could leverage our insights and incorporate behavioral priors, be fine-tuned, or be directly trained with task-dependent gaze data. Finally, many virtual experiences are carefully designed with specific expectations for user behavior, such as narrative experiences [57], architectural design [19], or training and education applications [45]. However, in contrast to traditional 2D displays, users have full control of the camera, and their actions, partly driven by their visual behavior, may not align with the content creator's expectations. Our analysis reveals significant differences in visual behavior between tasks, which content creators could consider when designing mechanisms such as audiovisual cues or virtual assistants to subtly direct users towards a desired task.

Our study investigates how visual behavior varies across different tasks in 3D immersive environments using a within-subjects design. However, further research is needed to understand how visual behavior changes when more complex cognitive processes are involved. For instance, social environments usually require multimodal interaction and social behaviors. Exploring these effects remains an interesting avenue for future research.

We have focused on visual behavior, which is widely used in similar studies. Future research could explore the potential benefits of integrating other physiological measurements such as galvanic skin response or heart rate, in order to gain a more comprehensive understanding of the users' cognitive and affective state.

Our analysis shows that the patterns in visual behavior we identified are robust across participants and stimuli, nevertheless, future studies may benefit from recruiting participants from a wider range of backgrounds and demographics or exploring more complex scenarios. This could uncover new insights and provide a more comprehensive understanding of how cognitive tasks affect visual behavior.

We hope that our study and publicly available data inspires further research in the aforementioned or other directions.

Acknowledgments

We extend our gratitude to the participants of our user studies. We would also like to thank the anonymous reviewers for their insightful comments. Our work has received funding from the European Union's Horizon 2020 research and innovation programme (ERC project CHAMELEON, Grant No 682080, and Marie Sklodowska-Curie project PRIME, Grant No 956585). This project was also funded by the Spanish Agencia Estatal de Investigación (projects PID2019-105004GB-I00 and PID2022-141539NB-I00). Additionally, Sandra Malpica and Daniel Martin were supported by a Gobierno de Aragon predoctoral grant (2018-2022 and 2020–2024 respectively).

References

[1] D. Abeles, R. Amit, and S. Yuval-Greenberg. Oculomotor behavior during non-visual tasks: The role of visual saliency. *PLOS One*, 13(6), 2018. 1

[2] M. Assens Reina, X. Giro-i Nieto, K. McGuinness, and N. E. O'Connor. Saltinet: Scan-path prediction on 360 degree images using saliency volumes. In *Proceedings of Computer Vision and Pattern Recognition (CVPR) Workshops*, pp. 2331–2338, 2017. 2

[3] D. H. Ballard, M. M. Hayhoe, and J. B. Pelz. Memory representations in natural tasks. *Journal of Cognitive Neuroscience*, 7(1):66–80, 1995. 1

[4] J. Battista, M. Kalloniatis, and A. Metha. Visual function: the problem with eccentricity. *Clinical and Experimental Optometry*, 88(5):313–321, 2005. 7

[5] T. Berger and M. Raschke. Repetition effects in task-driven eye movement analyses after longer time-spans. In *ACM Symposium on Eye Tracking Research and Applications*, pp. 1–6, 2021. 2, 3

[6] E. Bernal-Berdun, D. Martin, D. Gutierrez, and B. Masia. Sst-sal: A spherical spatio-temporal approach for saliency prediction in 360º videos. *Computers & Graphics*, 106, 2022. 2, 8

[7] B. M. Bolker, M. E. Brooks, C. J. Clark, S. W. Geange, J. R. Poulsen, M. H. H. Stevens, and J.-S. S. White. Generalized linear mixed models: a practical guide for ecology and evolution. *Trends in Ecology & Evolution*, 24(3):127–135, 2009. 4

[8] C. Bryan, A. Mishra, H. Shidara, and K.-L. Ma. Analyzing gaze behavior for text-embellished narrative visualizations under different task scenarios. *Visual Informatics*, 4(3):41–50, 2020. 2

[9] F.-Y. Chao, C. Ozcinar, C. Wang, E. Zerman, L. Zhang, W. Hamidouche, O. Deforges, and A. Smolic. Audio-visual perception of omnidirectional video for virtual reality applications. In *International Conference on Multimedia & Expo Workshops (ICMEW)*, pp. 1–6, 2020. 2, 4

[10] F.-Y. Chao, C. Ozcinar, L. Zhang, W. Hamidouche, O. Deforges, and A. Smolic. Towards audio-visual saliency prediction for omnidirectional video with spatial audio. In *International Conference on Visual Communications and Image Processing (VCIP)*, pp. 355–358, 2020. 2

[11] E. S. Dalmaijer, S. Mathôt, and S. Van der Stigchel. Pygaze: An open-source, cross-platform toolbox for minimal-effort programming of eye-tracking experiments. *Behavior Research Methods*, 46:913–921, 2014. 3

[12] D. Draschkow, J. M. Wolfe, and M. L.-H. Vo. Seek and you shall remember: Scene semantics interact with visual search to build better memories. *Journal of Vision*, 14(8):10–10, 2014. 1

[13] L. R. Enders, R. J. Smith, S. M. Gordon, A. J. Ries, and J. Touryan. Gaze behavior during navigation and visual search of an open-world virtual environment. *Frontiers in Psychology*, 12:681042, 2021. 2, 3

[14] J. Fawcett, E. Risko, and A. Kingstone. *The handbook of attention*. MIT Press, 2015. 1

[15] J. R. Flanagan, Y. Terao, and R. S. Johansson. Gaze behavior when reaching to remembered targets. *Journal of Neurophysiology*, 100(3):1533–1543, 2008. 2

[16] J. O. Goh, J. C. Tan, and D. C. Park. Culture modulates eye-movements to visual novelty. *PLOS One*, 4(12):e8238, 2009. 3

[17] J. Hadnett-Hunter, G. Nicolaou, E. O'Neill, and M. Proulx. The effect of task on visual attention in interactive virtual environments. *ACM Transactions on Applied Perception (TAP)*, 16(3):1–17, 2019. 2, 7, 8

[18] D. E. Hannula, R. R. Althoff, D. E. Warren, L. Riggs, N. J. Cohen, and J. D. Ryan. Worth a glance: using eye movements to investigate the cognitive neuroscience of memory. *Frontiers in Human Neuroscience*, 4:166, 2010. 2

[19] J. Haskins, B. Zhu, S. Gainer, W. Huse, S. Eadara, B. Boyd, C. Laird, J. Farantatos, and J. Jerald. Exploring vr training for first responders. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops*, pp. 57–62. IEEE, 2020. 8

[20] M. Hayhoe and D. Ballard. Eye movements in natural behavior. *Trends in Cognitive Sciences*, 9(4):188–194, 2005. 1

[21] M. M. Hayhoe, T. McKinney, K. Chajka, and J. B. Pelz. Predictive eye movements in natural vision. *Experimental Brain Research*, 217(1):125–136, 2012. 1

[22] M. M. Hayhoe, A. Shrivastava, R. Mruczek, and J. B. Pelz. Visual memory and motor planning in a natural task. *Journal of Vision*, 3(1):6–6, 2003. 1

[23] Z. Hu. Gaze analysis and prediction in virtual reality. In *IEEE Conference on Virtual Reality and 3D User Interfaces Abstracts and Workshops*, pp. 543–544, 2020. 7

[24] Z. Hu, A. Bulling, S. Li, and G. Wang. Ehtask: Recognizing user tasks from eye and head movements in immersive virtual reality. *IEEE Transactions on Visualization and Computer Graphics*, 29(4), 2021. 2, 7, 8

[25] H. Huang, N.-C. Lin, L. Barrett, D. Springer, H.-C. Wang, M. Pomplun, and L.-F. Yu. Analyzing visual attention via virtual environments. In *SIGGRAPH ASIA 2016 Virtual Reality meets Physical Reality: Modelling and Simulating Virtual Humans and Environments*, pp. 1–2. 2016. 1

[26] L. Itti and C. Koch. Computational modelling of visual attention. *Nature Reviews Neuroscience*, 2(3):194–203, 2001. 1

[27] X. Jin, J. Meneely, and N.-K. Park. Virtual reality versus real–world space: Comparing perceptions of brightness, glare, spaciousness, and visual acuity. *Journal of Interior Design*, 47(2):31–50, 2022. 7

[28] B. John, P. Raiturkar, O. Le Meur, and E. Jain. A benchmark of four methods for generating 360º saliency maps from eye tracking data. *International Journal of Semantic Computing*, 13(03):329–341, 2019. 8

[29] E. Johnstone and C. A. Dodd. Placements as mediators of brand salience within a uk cinema audience. *Journal of Marketing Communications*, 6(3):141–158, 2000. 1

[30] J. K. Kaakinen and J. Hyönä. Perspective effects in repeated reading: An eye movement study. *Memory & cognition*, 35:1323–1336, 2007. 3

[31] A. Kafkas and D. Montaldi. Recognition memory strength is predicted by pupillary responses at encoding while fixation patterns distinguish recollection from familiarity. *Quarterly Journal of Experimental Psychology*, 64(10):1971–1989, 2011. 2

[32] J. E. Kamienkowski, M. J. Carbajal, B. Bianchi, M. Sigman, and D. E. Shalom. Cumulative repetition effects across multiple readings of a word: Evidence from eye movements. *Discourse Processes*, 55(3):256–271, 2018. 3

[33] D. Kit, L. Katz, B. Sullivan, K. Snyder, D. Ballard, and M. Hayhoe. Eye movements, visual search and scene memory, in an immersive virtual environment. *PLOS One*, 9(4):e94362, 2014. 2

[34] R. Kothari, Z. Yang, C. Kanan, R. Bailey, J. B. Pelz, and G. J. Diaz. Gaze-in-wild: A dataset for studying eye and head coordination in everyday activities. *Scientific Reports*, 10(1):2539, 2020. 7

[35] G.-A. Koulieris, B. Bui, M. S. Banks, and G. Drettakis. Accommodation and comfort in head-mounted displays. *ACM Transactions on Graphics (TOG)*, 36(4):1–11, 2017. 7

[36] J. Kurz, M. Hegele, and J. Munzert. Gaze behavior in a natural environment with a task-relevant distractor: How the presence of a goalkeeper distracts the penalty taker. *Frontiers in Psychology*, 9:19, 2018. 1

[37] M. F. Land and M. Hayhoe. In what ways do eye movements contribute to everyday activities? *Vision Research*, 41(25-26):3559–3565, 2001. 1

[38] M. F. Land and D. N. Lee. Where we look when we steer. *Nature*, 369(6483):742–744, 1994. 1

[39] O. Le Meur and Z. Liu. Saccadic model of eye movements for free-viewing condition. *Vision Research*, 116:152–164, 2015. 1, 7

[40] C.-L. Li, M. P. Aivar, D. M. Kit, M. H. Tong, and M. M. Hayhoe. Memory and visual search in naturalistic 2D and 3D environments. *Journal of Vision*, 16(8):9–9, 2016. 1, 2

[41] H.-I. Liao and S. Shimojo. Dynamic preference formation via gaze and memory. In *Neuroscience of Preference and Choice*, pp. 277–292. 2012. 1

[42] G. Llorach, A. Evans, and J. Blat. Simulator sickness and presence using HMDs: comparing use of a game controller and a position estimation system. In *Proceedings of the 20th ACM Symposium on Virtual Reality Software and Technology*, pp. 137–140, 2014. 7

[43] C. Marañes, D. Gutierrez, and A. Serrano. Exploring the impact of 360° movie cuts in users' attention. In *IEEE Conference on Virtual Reality and 3D User Interfaces (IEEE VR)*, 2020. 2

[44] D. Martin, D. Gutierrez, and B. Masia. A probabilistic time-evolving approach to scanpath prediction. *ArXiV (Preprint)*, 2022. 1

[45] D. Martin, S. Malpica, D. Gutierrez, B. Masia, and A. Serrano. Multimodality in vr: A survey. *ACM Computing Surveys*, 54(10):1–36, 2022. 8

[46] D. Martin, A. Serrano, A. W. Bergman, G. Wetzstein, and B. Masia. Scangan360: A generative model of realistic scanpaths for 360 images. *IEEE Transactions on Visualization and Computer Graphics*, 28(5):2003–2013, 2022. 2, 8

[47] D. Martin, A. Serrano, and B. Masia. Panoramic convolutions for 360° single-image saliency prediction. In *CVPR Workshop on Computer Vision for Augmented and Virtual Reality*, vol. 2, 2020. 2

[48] Matlab. Statistics and Machine Learning Toolbox. `https://www.mathworks.com/products/statistics.html`. Last Accessed: 2023-07-25. 4

[49] A. Milisavljevic, T. L. Bras, M. Mancas, C. Petermann, B. Gosselin, and K. Doré-Mazars. Towards a better description of visual exploration through temporal dynamic of ambient and focal modes. In *Proceedings of ACM Symposium on Eye Tracking Research & Applications*, pp. 1–4, 2019. 7

[50] T. Mustonen, M. Berg, J. Kaistinen, T. Kawai, and J. Häkkinen. Visual task performance using a monocular see-through head-mounted display (HMD) while walking. *Journal of Experimental Psychology: Applied*, 19(4):333, 2013. 2

[51] M. B. Neider and G. J. Zelinsky. Scene context guides eye movements during visual search. *Vision Research*, 46(5):614–621, 2006. 2

[52] C. Ozcinar and A. Smolic. Visual attention in omnidirectional video for virtual reality applications. In *International Conference on Quality of Multimedia Experience (QoMEX)*, pp. 1–6, 2018. 1

[53] M. Pejić, G. Savić, and M. Segedinac. Determining gaze behavior patterns in on-screen testing. *Journal of Educational Computing Research*, 59(5):896–925, 2021. 1

[54] Y. Rai, J. Gutiérrez, and P. Le Callet. A dataset of head and eye movements for 360 degree images. In *Proceedings of ACM Multimedia Systems Conference*, pp. 205–210, 2017. 2

[55] S. L. Rogers, C. P. Speelman, O. Guidetti, and M. Longmuir. Using dual eye tracking to uncover personal gaze patterns during social interaction. *Scientific Reports*, 8(1):1–9, 2018. 1, 5

[56] A. C. Schütz, D. I. Braun, and K. R. Gegenfurtner. Eye movements and perception: A selective review. *Journal of Vision*, 11(5):9–9, 2011. 1

[57] A. Serrano, V. Sitzmann, J. Ruiz-Borau, G. Wetzstein, D. Gutierrez, and B. Masia. Movie editing and cognitive event segmentation in virtual reality video. *ACM Transactions on Graphics*, 36(4):1–12, 2017. 2, 8

[58] V. Sitzmann, A. Serrano, A. Pavel, M. Agrawala, D. Gutierrez, B. Masia, and G. Wetzstein. How do people explore virtual environments? *IEEE Transactions on Visualization and Computer Graphics*, 24(4):1633–1642, 2017. 2, 3, 6, 7

[59] B. Sullivan, C. Rothkopf, M. Hayhoe, and D. Ballard. Task-dependent gaze priorities in driving. *Journal of Vision*, 11(11):932–932, 2011. 1

[60] W. Sun, Z. Chen, and F. Wu. Visual scanpath prediction using ior-roi recurrent mixture density network. *IEEE Transactions on Pattern Analysis and Machine Intelligence*, 43(6):2101–2118, 2019. 1

[61] B. V. Syiem, R. M. Kelly, J. Goncalves, E. Velloso, and T. Dingler. Impact of task on attentional tunneling in handheld augmented reality. In *Proceedings of the 2021 CHI Conference on Human Factors in Computing Systems*, pp. 1–14, 2021. 2

[62] P. M. van Leeuwen, R. Happee, and J. C. de Winter. Changes of driving performance and gaze behavior of novice drivers during a 30-min simulator-based training. *Procedia Manufacturing*, 3:3325–3332, 2015. 2

[63] G. F. Woodman and S. J. Luck. Serial deployment of attention during visual search. *Journal of Experimental Psychology: Human Perception and Performance*, 29(1):121, 2003. 7

[64] R. Xu. Measuring explained variation in linear mixed effects models. *Statistics in Medicine*, 22(22):3527–3541, 2003. 6

[65] Y. Zhang. Asod60k: An audio-induced salient object detection dataset for panoramic videos. *ArXiV (Preprint)*, 2021. 4

[66] D. Zhu, X. Shao, Q. Zhou, X. Min, G. Zhai, and X. Yang. A novel lightweight audio-visual saliency model for videos. *ACM Transactions on Multimedia Computing, Communications and Applications*, 19(4), 2022. 2

[67] G. Ziv. Gaze behavior and visual attention: A review of eye tracking studies in aviation. *The International Journal of Aviation Psychology*, 26(3-4):75–104, 2016. 2