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

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

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.

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.
behavior. We found that, even though head movements vary significantly between tasks, eye eccentricity is not significantly affected by 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 360° 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 behavior 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.

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 variability 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 how 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 (CEIC). 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].
We 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.

<table>
<thead>
<tr>
<th>Variable</th>
<th>Order</th>
<th>t-stat</th>
<th>p-value</th>
<th>Variation</th>
<th>t-stat</th>
<th>p-value</th>
<th>Order:Variation</th>
<th>t-stat</th>
<th>p-value</th>
</tr>
</thead>
<tbody>
<tr>
<td>Number of fixations</td>
<td>0.248</td>
<td>0.806</td>
<td>0.925</td>
<td>-0.508</td>
<td>0.616</td>
<td>0.925</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration of fixations</td>
<td>0.093</td>
<td>0.925</td>
<td>0.925</td>
<td>0.066</td>
<td>0.947</td>
<td>0.947</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Dwell time</td>
<td>0.001</td>
<td>0.999</td>
<td>0.999</td>
<td>-1.077</td>
<td>0.303</td>
<td>0.303</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Number of saccades</td>
<td>0.163</td>
<td>0.870</td>
<td>0.870</td>
<td>1.057</td>
<td>0.290</td>
<td>0.290</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Duration of saccades</td>
<td>1.540</td>
<td>0.123</td>
<td>0.123</td>
<td>-0.064</td>
<td>0.948</td>
<td>0.948</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Amplitude of saccades</td>
<td>0.068</td>
<td>0.945</td>
<td>0.945</td>
<td>0.542</td>
<td>0.590</td>
<td>0.590</td>
<td></td>
<td></td>
<td></td>
</tr>
<tr>
<td>Eye eccentricity</td>
<td>-0.408</td>
<td>0.685</td>
<td>0.685</td>
<td>-0.102</td>
<td>0.918</td>
<td>0.918</td>
<td></td>
<td></td>
<td></td>
</tr>
</tbody>
</table>

**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 $\times$ 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) $\times$ 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
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 \(\Omega^2_p\) which is suitable for a non-parametric analysis like the GLMM. We consider small effect sizes \(\Omega^2_p < 0.01\), medium effect sizes \(0.01 \leq \Omega^2_p \leq 0.06\) and big effect sizes \(\Omega^2_p > 0.06\), and establish significance level at \(p = 0.05\). Post-hoc analyses were carried out via marginal means with the `emmeans` Matlab package.

### 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 oculomotor 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 ($t = 6.393$, $p < 0.001$, $\Omega_p^2 = 0.16$). 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 to the same visual information, performing more, larger movements in a shorter time, with the head. 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 parfoveal 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 Skłodowska-Curie project PRIME, Grant No 956585). This project was also funded by the Spanish Agencia Estatal de Investigación (projects PID2019-105004GB-100 and PID2022-141539NB-100). Additionally, Sandra Malpica and Daniel Martin were supported by a Gobierno de Aragon predocctoral grant (2018-2022 and 2020–2024 respectively).

REFERENCES


