# Attentive Saliency and Photorealism in Immersive Virtual Environments


Deborah Pintani*
Department of Computer Science
University of Verona
Verona, Italy
deborah.pintani@univr.it

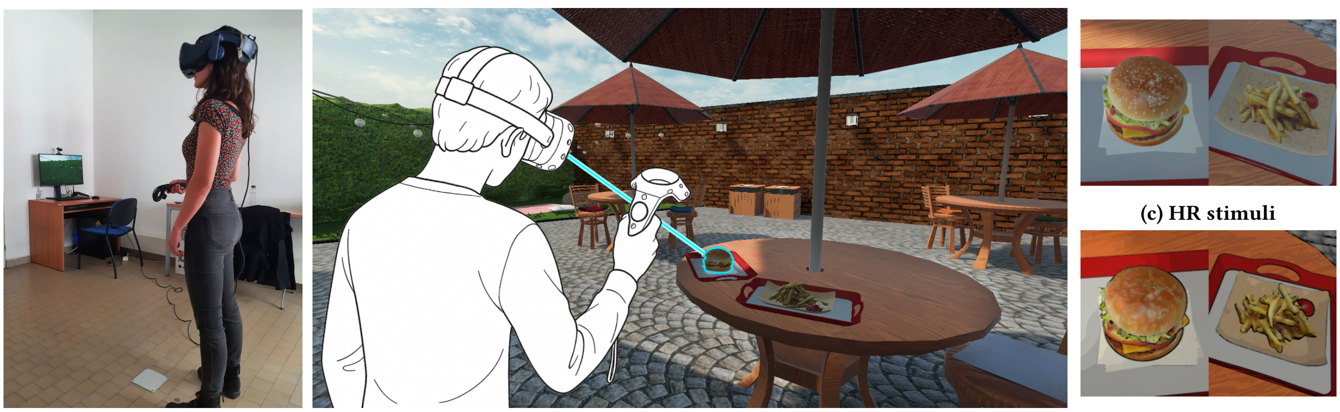
Giulia Benvegnù
Department of Diagnostic and Public Health
University of Verona
Verona, Italy
giulia.benvegnu@univr.it

Federico Maria Lorusso
Department of Diagnostic and Public Health
University of Verona
Verona, Italy
federicomaria.lorusso@studenti.unipd.it

Cristiano Chiamulera
Department of Diagnostic and Public Health
University of Verona
Verona, Italy
cristiano.chiamulera@univr.it

Andrea Giachetti
Department of Engineering for Innovation Medicine
University of Verona
Verona, Italy
andrea.giachetti@univr.it

Ariel Caputo
Department of Engineering for Innovation Medicine
University of Verona
Verona, Italy
ariel.caputo@univr.it

(a) Setup (b) User task (c) HR stimuli (d) LR stimuli

**Figure 1: Participants of our experiments (a) freely navigate in environments (b) rendered with two different degrees of realism. We study the attentive saliency of food-related stimuli in the High Realism setting (c) and in the Low Realism setting (d)**


## Abstract

Virtual Reality is widely used in behavioral studies, as it allows to simulate realistic experiences in environments completely controlled by the study designers. In particular, studies on visual saliency and attention can leverage immersive graphics rendered on head-mounted displays to create custom environments and head, body, and gaze tracking systems to evaluate the participant's behavior. However, some aspects of the design of VR applications for attentive saliency study, especially in relation to rendering style, are poorly investigated in the literature. In our work, we created virtual environments for a study on junk food craving, and performed an experiment to assess how much the saliency of visual items is influenced by photorealism and correlated with the presence measured with questionnaires. Furthermore, we investigated the feasibility of replacing the eye-tracking-based saliency measure with approximated measurements based on head tracking.




*CHItaly 2025, Salerno, Italy*



ACM ISBN 979-8-4007-2102-1/25/10
https://doi.org/10.1145/3750069.3750321



## CCS Concepts

• **Human-centered computing** → **Virtual reality**; **User studies**;
• **Computing methodologies** → **Virtual reality**.

## Keywords

Virtual Reality, Attentive Saliency, Visual Attention, Rendering Style, Gaze Tracking

**ACM Reference Format:**
Deborah Pintani, Giulia Benvegnù, Federico Maria Lorusso, Cristiano Chiamulera, Andrea Giachetti, and Ariel Caputo. 2025. Attentive Saliency and Photorealism in Immersive Virtual Environments. In *CHItaly 2025: 16th Biannual Conference of the Italian SIGCHI Chapter (CHItaly 2025), October 06–10, 2025, Salerno, Italy*. ACM, New York, NY, USA, 8 pages. https://doi.org/10.1145/3750069.3750321

# 1 Introduction

Immersive Virtual Environments (IVEs) are widely employed to perform behavioral studies, as they allow for complete control of the environment and an automatic estimation of behavioral measurements. This is clearly due to the fact that Virtual Reality technology is now providing a high level of presence [2], but also to the use of head/hand tracking systems and, in some cases, eye-tracking systems integrated in the headset [1] that can automatically record data on user actions and attention.

IVEs have been used to stimulate emotional reactions using specific visual cues to study or treat food craving [12], alcohol or smoking addictions [14, 25], or shopping behavior [5].

While this huge amount of work demonstrates the potential impact of Virtual Reality (VR) tools on behavioral research, some technical aspects related to the design of these studies appear not fully investigated, which could limit the robustness and generalizability of research findings.

In this paper, we exploit IVEs created for research on food craving, to investigate two specific choices that, in our opinion, need to be better understood. The first is related to the degree of graphical realism of the IVE. While many studies investigated the correlation between the level of graphical realism and the sense of presence [8, 13, 16] (with not always consistent results), it is not clear how attentive tasks are influenced by the different levels of photorealism. The first research question we aim to address is thus the following (RQ1): "Does an increased photorealism of the scene change the saliency of attentive stimuli?".

The other question tackled by our study is the necessity of using eye-tracking systems to measure attentive saliency in this kind of research works. While some Head-Mounted Displays, like the VIVE Eye Pro or Microsoft HoloLens 2, integrate eye-tracking systems, it has been demonstrated that attention maps on scene elements can be approximated quite well with the simple tracking of head position and orientation [23]. The question here is the following (RQ2): "Is there a difference in the evaluation of attentive behavior performed with gaze tracking system or with head-orientation based measurement?".

To answer these questions, we analyzed the outcomes of a specifically designed experiment, where we measured how much food-related stimuli capture the attention of human participants immersed in Virtual Environments rendered with different levels of photorealism, estimating the saliency of the stimuli employing gaze-based ray casting or head-based ray casting. The experiment outcomes provide useful insights for designing novel applications to study attentive saliency using Virtual Reality.

The paper is organized as follows: Section 2 gives a background on the use of VR for studying attention and on the previous works analyzing the effects of varying photorealism in VR. Section 3 provides details on the experimental design and on the technical solutions adopted. Section 4 presents the experimental results, which are discussed in Section 5.

The code and materials related to this study are publicly available at https://github.com/univr-GRAIL/VR_Saliency-Photorealism.

# 2 Related Work

## 2.1 VR for visual saliency assessment

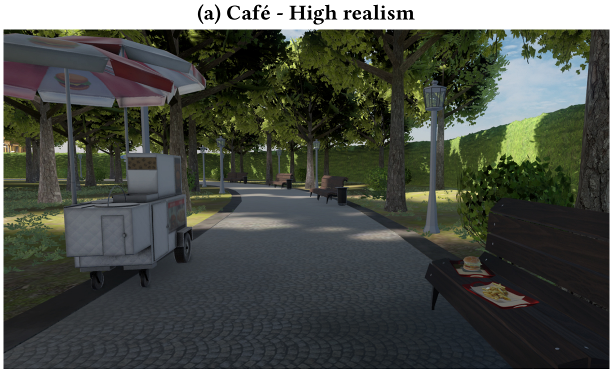
In recent years, the introduction of VR and Augmented Reality (AR) as more mainstream consumer products led to the integration of more advanced features, such as eye-tracking sensors, in a larger number of devices. Researchers have been increasingly using HMD models with eye-tracking capabilities to study users' behavior and cognitive reactions, such as attention analysis, saliency mapping, gaze-based interaction models, and so on. A potential advantage of performing these studies using IVEs and eye-tracking technology is the ability for researchers to recreate specific scenarios. Furthermore, compared to real-world settings, an IVE provides fewer uncontrollable variables that can interfere with the experiments. On this topic, Benvegnù et al. [3] conducted a study to investigate the suitability of VR scenarios on approaches such as Environmental Enrichment (EE), aimed at reducing food craving behaviors in healthy volunteers. The authors tested the EE on both real-world and VR conditions and found that the effect in the latter was not only present but also amplified. This shows how IVEs inevitably introduce their own set of factors that differ from the real world. For example, the level of immersion, the virtual stimuli and interacting with a non-physical environment can influence the user in a different way. The effect of these factors is the object of study in several works in recent literature. Olk et al. [17] performed a study to measure attention and distraction in IVEs, finding that traditional paradigms can be translated and applied in VR. However, the two conditions (i.e., reality and virtual reality) present some differences, as shown by Biermeier et al. [4]. The way behaviors, like attention, are measured during experiments is another factor that can depend on the technology implemented. For example, Llanes-Jurado et al. [15] studied the relationship between the area of interest featured in a virtual scenario and the method used to track the user's attention (i.e., eye tracking vs head direction). Furthermore, IVEs' characteristics can affect studies on behavior and attention, depending on the visual features of the virtual scenarios the user experiences during the experiments. In particular, graphics aspects such as rendering realism, screen and texture resolution represent new factors in studies that can influence experimental outcomes in different ways. Gonçalves et al. [10] performed a review of comparative studies of the impact of realism in IVEs, reporting that most of the studies found that a higher realism correlates with a better perceived experience for the users. However, they acknowledge that there are more realism-defining factors and their impact on specific parts of the user experience should be further studied.

## 2.2 Photorealism and presence

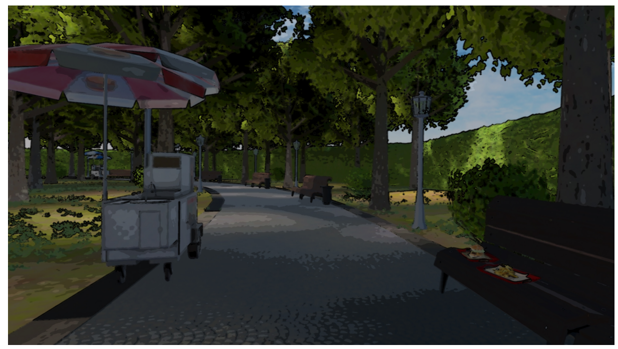
A different category of studies also highlights the importance that photorealism in IVEs has on the perceived users' feeling of presence. For example, Newman et al. [16] performed a study that tries to assess whether relevant benefits in terms of user experience justify the cost of expensive equipment needed to achieve a high realism degree. They measured users' responses testing different factors such as the level of photorealism, using respectively high and low-poly meshes for the 3D models, and the type of environment (i.e. natural, featuring a park-like environment vs. built, featuring a more urban setting). They found that high photorealism strongly affects the


**(a) Café - High realism**


**(b) Café - Low realism**



**(c) Park - High realism**



**(d) Park - Low realism**

**Figure 2: The two experimental scenes (Café and Park) shown in both High Realism and Low Realism versions. For a clearer view of the differences in rendering style, please refer to Figure 5**

sense of presence. Another study [19] comparing different levels of photorealism using different lighting methods (i.e. ray-casting vs. real-time ray-tracing), showed another aspect of photorealism that can affect the perceived presence. A similar work from Gonçalves et al. [11] further investigates aspects of the virtual scene lighting by considering new factors such as global illumination, screen-space reflections, ambient occlusion and direct shadow casting. In this case, the authors found that some factors play a bigger role than others in the perception of the scene realism, showing the complexity behind the task of measuring presence. DeMarbre et al. [8] examined how the ratio of virtual to real content affects the sense of presence. In addition to this variable, they also controlled the visual style, comparing realistic and stylized environments. Their results showed that users preferred realistic virtual content, resulting in a higher sense of presence than the stylized version. A wide review of works tackling this topic is found in a rather extensive survey from Souza et al. [22].

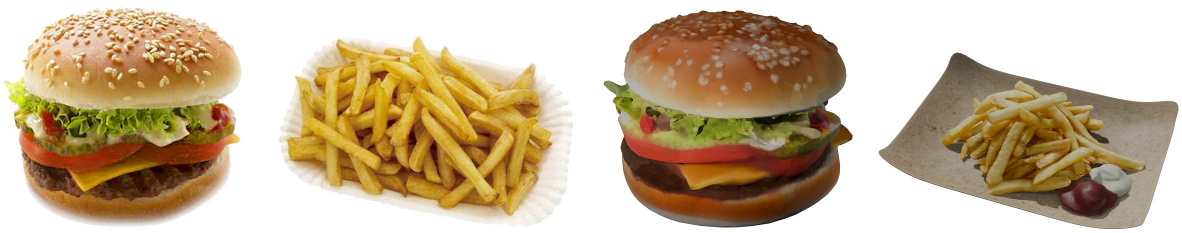
## 3 Experiment design and setup

### 3.1 User study

To answer our research questions, we designed an experiment using immersive virtual environments where participants can freely navigate, see, and manipulate 3D models of junk food, as shown in Figure 1. The experiment was approved by the Person Research Approval Committee (CARP) of the University of Verona. To tackle RQ1, we introduced a photorealism factor by building two outdoor scenes (i.e., a café and a park setting), and from each of them, we created two versions with different degrees of photorealism: High Realism (HR) and Low Realism (LR), as shown in Figure 2. The experiment took place in a laboratory room with a VR setup consisting of an HTC Vive Pro Eye HMD, a PC equipped with an i7-14700KF Intel CPU and an RTX 4070 Nvidia GPU used to run the VR application, and a delimited zone of 2.65 by 3.6 meters dedicated to VR interaction. Before starting the experiment, we collected information about age, gender, experience with VR applications, videogames, and relevant eye conditions. Using the HMD, each user performed an initial eye calibration and experienced both the café and park scenarios for 4 minutes, in either the HR or LR condition. In both scenarios, users were tasked to freely explore the environment, and informed of the possibility to interact with certain designated objects. To avoid order effects, half the participants started with the park scenario and the others with the café scenario. During the time of each run, the application recorded data, collecting several measurements related to the attention. In particular, the HMD's eye-tracker sensor was used to record the user's gaze direction, and the tracking of the HMD in the real world was used to record the direction in which the user's head was aiming in the virtual environment. Additionally, we recorded information about the specific interactable object that was the target of either the user's gaze or the head direction.


**(a) Teleporting** **(b) Grabbing** **(c) Controller mapping**

**Figure 3: User actions supported in the Virtual Environments (a) navigation with teleporting. (b) object grabbing. (c) Buttons used for the actions in the VIVE controller.**



**(a)** **(b)** **(c)** **(d)**

**Figure 4: (a),(b): validated and standardized food-related stimuli from [9]. (c),(d): renderings of the corresponding 3D models inserted in our environments.**

Users waited a minimum of 3 minutes before starting the second run and, at the end of both runs, users were asked to fill out the Presence Questionnaire (PQ) [26]. The questionnaire consists of 29 questions used to measure the user's presence in the virtual scenario and provides a final score divided into 4 categories: *involvement*, *interface quality*, *adaptation/immersion* and *visual fidelity*.

## 3.2 Virtual environments

We created two different virtual scenes to support our research questions: a natural environment (Park), and an urban setting (Café). The different contexts have been designed to support other future research dedicated to studying the influence of environmental characteristics on addictive behaviors.

The VR interactive environments are developed with Unity 2022.3.0, using SteamVR [20] (which implements OpenVR APIs [21]). The head-mounted display (HMD) of choice is the HTC VIVE Pro Eye, as it features a built-in eye-tracking system. We use SteamVR Input to manage inputs coming from Vive Pro Controllers, and Auto-Hand plugin [18] for physics interactions.

In the applications, the participants need to be able to freely navigate in the environments and grab special objects of interest.

Navigation is supported with a teleport method, where the user selects the direction by pointing the controller holding down the Trackpad button (blue in Figure 3c). A green circle and an arc appear if the location allows teleporting (if not, the circle and the arc will be red.). The user can then release the button to be automatically teleported there or squeeze the Grip button (green in Figure 3c) to cancel the intended action.

To grab an interactable, players need to approach it and hold down the Trigger button (red in Figure 3c). Releasing the button, they can drop the item. When the Trigger button is pushed, a white circle appears on top of the controller.

The Unity application automatically records a wide range of system measurements, namely the player's initial position in the virtual world, which is used as a reference for all other relative positions and a continuous series of values sampled every 0.2 seconds: system time, head position and orientation.

For the purpose of the study, we defined two *special* objects, both belonging to the junk food category and known for being highly palatable and appetizing: a Burger and a paper tray of French Fries. These items are considered *special* because they are interactable (i.e., users can grab them) and are the only objects monitored for user attention, specifically through gaze and head pointing measurements. Both the burger and the tray of fries are treated as single, unified objects: users can grab the entire item (e.g., the whole tray of fries), but not individual components (e.g., individual fries).

Both food cues are taken from the FoodCast Research Image Database (FRIDa) [9] to ensure the use of validated and standardized food-related stimuli. After selecting images from FRIDa that matched our chosen stimuli, we created similar 3D models based on these references. Figure 4 depicts the original FRIDa images and their derived 3D models.

## 3.3 Measurement of attentive salience

To study the attentive salience of the *special* objects, we rely, as a first option, on the Tobii eye-tracking sensor integrated in the headset. Using both SRanipal [6] and Tobii XR SDK [24], it is possible to obtain the eye gaze direction represented as a ray cast, and an event is recorded whenever a *special* object intersects the ray. In our application, we record the gaze-based focus (GBF) on the *special* objects as a binary flag (on/off) every 0.2 seconds.

To investigate the possible realization of a VR attention study without an eye-tracker software and to answer our RQ2, we implemented an alternative way to detect attention on *special* objects based on head orientation.

Head pointing is implemented as a sphere cast from the player's head, represented by the *Camera* object in the virtual world. The sphere has a radius of 10 cm (diameter 20 cm), corresponding to the physical size of the headset's front. When the first object hit by the cast is a *special* item, the event is recorded.

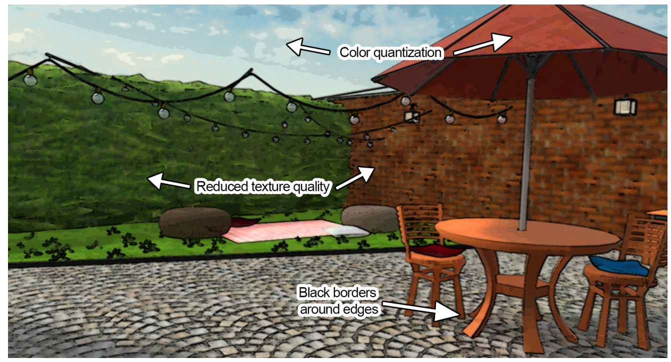
The use of a margin in the ray-casting procedure is motivated by the idea of considering a tolerance in the correspondence of the directions tracked by head and eye trackers, which has a similar order of magnitude as reported in [15]. In our application, we record the head-based focus (HBF) on *special* objects as a binary flag (on/off) every 0.2 seconds.

## 3.4 Implementation of the different levels of realism

Each scene is created in two variants: a High Realism (HR) version and a Low Realism (LR) version. Both variants are rendered using Unity's Universal Render Pipeline (URP). As there are too many parameters that can affect photorealism in 3D graphics rendering, we decided not to control single parameters individually, but to produce only two versions of the scenes: one including high quality textures and tricks to increase photorealism (HR) and the other reducing the realism of the effects by creating a cartoon-like rendering (LR).

In detail, for the HR version, we used high-quality 2K textures and high-quality soft shadows settings. High-resolution soft shadows better simulate the behavior of real lights, with smoother and less sharp edges (Figures 2a, 2c, 5b). This improves the photorealism of the environment, making it more believable. Additionally, after building the application, we used ReShade [7] to post-process frames. We simulate High-Dynamic Range (HDR) that implements a tone mapper to better simulate outdoor highlights and shadows, preserving details in dark and bright areas. Moreover, we applied a bloom effect to create a glow around bright areas, making the lights appear more diffused and natural.

For the LR version, we decreased the texture resolution and the quality of shadows. Textures are reduced to $\frac{1}{8}$ of their original dimensions. To lower the realism, we only used hard shadows. Similarly to the HR version, we used ReShade to add post-process effects, but, in this case, to make the result less realistic. In particular, we simulated a cartoon style by adding: black borders around the edges, a slight surface blur, increased saturation by 20%, and a color quantization that limited the brightness values to 16 levels. These modifications resulted in a flatter and cartoonish style of rendering (Figures 2b, 2d, 5a). Both versions included baked lightmaps, ambient occlusion, and light probes in specific areas around the interactables' positions.



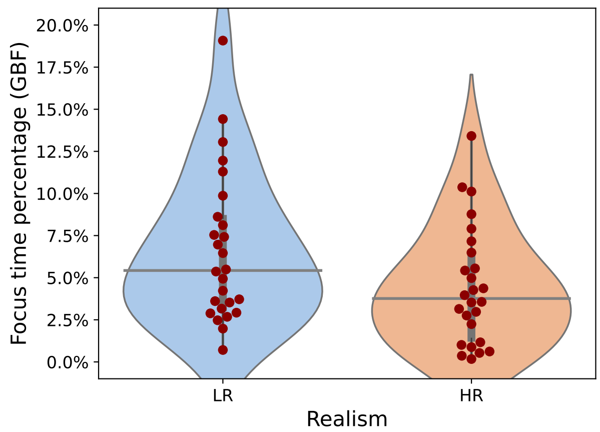

**(a) LR Café**

**(b) HR Café**

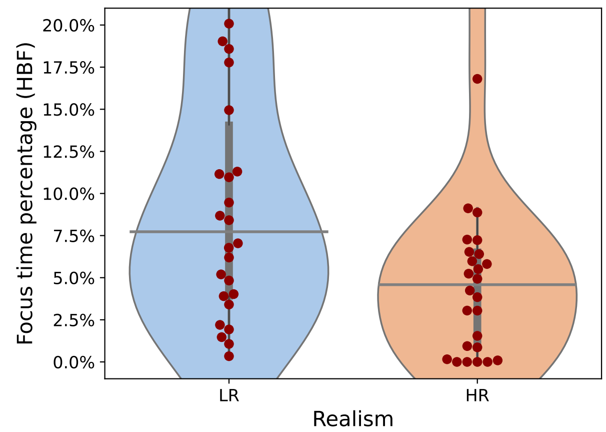
**Figure 5: Comparison of the same point of view in the Café scene. Arrows indicate parts of the image where effects are most noticeable.**

## 4 Experimental results

We conducted the experiment with 26 participants recruited from ages 19 to 44 years and an average of 26 years. Participants were divided equally between the two experimental conditions: LR = Low Realism and HR = High Realism, with each group consisting of 13 people. The HR group included 6 women and 7 men with a total average age of 26, while the LR group included 7 women and 6 men with a total average age of 25, providing a balanced age and gender distribution in both conditions. The two groups also featured the same number of people who stated to have a mid to high degree of expertise in the use of videogames (4 experts and 7 non-experts in each group) and, in both groups, we had no participants with particular expertise in the use of VR applications. Figure 6 shows the




(a)




(b)

**Figure 6: (a) Distributions of the percentages of the total experience time in which participants paid attention to the interactable items (burger and fries) for the two groups (LR=Low Realism, HR=High Realism), measured with the Gaze-Based Focus detection. (b) Same distributions measured with the Head-Based Focus detection.**

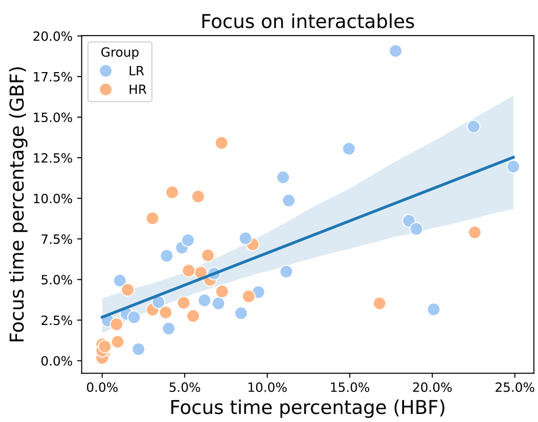



**Figure 7: Scatter plot representing gaze-based vs head-based saliency measured for special objects. The two measurements appear correlated. Only for a few participants, a low focus with one method corresponds to a high one measured with the other technique.**

percentages of the total experience time in which each participant paid attention to one of the interactable items (i.e., burger and fries) for the two groups LR and HR, measured with both the Gaze-Based method (a) and the Head-Based method (b). It is possible to see that the two methods provide consistent results and that there are non-negligible differences in the median values (horizontal lines) with a longer focus on items for the participants performing the experiment in the LR setting. The median values estimated with the GBF are 5.4% for the HR setting and 3.8% for the LR setting. The median values estimated with HBF are 4.6% for the HR setting and 7.7% for the LR setting. Data have been tested for normality using Shapiro-Wilk and resulted non-normally distributed ($p<0.001$). We then performed Man-Whitney U tests, and for the HBF method,

| Witmer's Presence Questionnaire Scores | | | | |
|---|---|---|---|---|
| | **Avg HR** | **SDev HR** | **Avg LR** | **SDev LR** |
| **Involvement** | 63,23 | 13,12 | 55,92 | 10,28 |
| **Sensory Fidelity** | 31,31 | 4,87 | 30,31 | 4,01 |
| **Immersion** | 45,85 | 7,03 | 44,62 | 6,28 |
| **Interface Quality** | 14,38 | 3,23 | 16,08 | 2,53 |
| **Total** | 154,77 | 23,72 | 146,92 | 16,77 |

**Table 1: Witmer's PQ scores [26] results per category and total. For each item, the average score and the standard deviation are reported.**

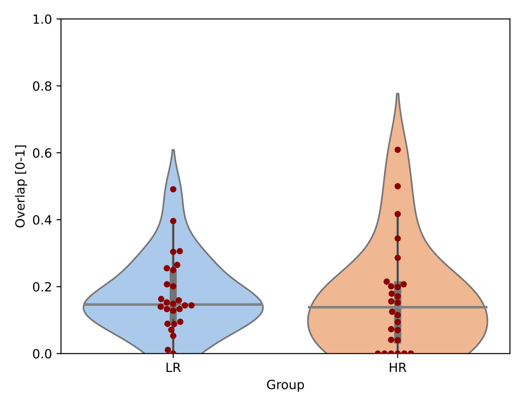
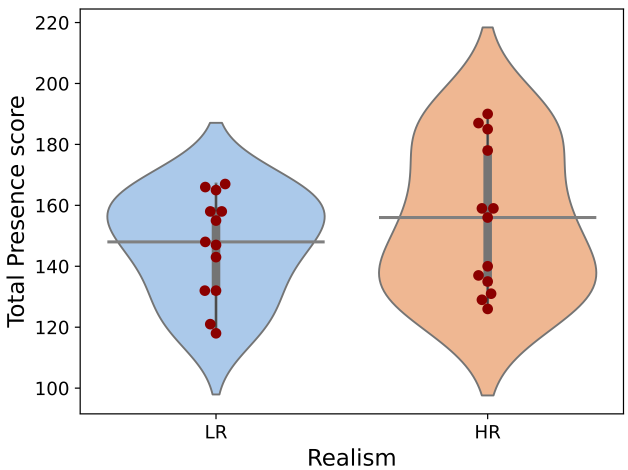


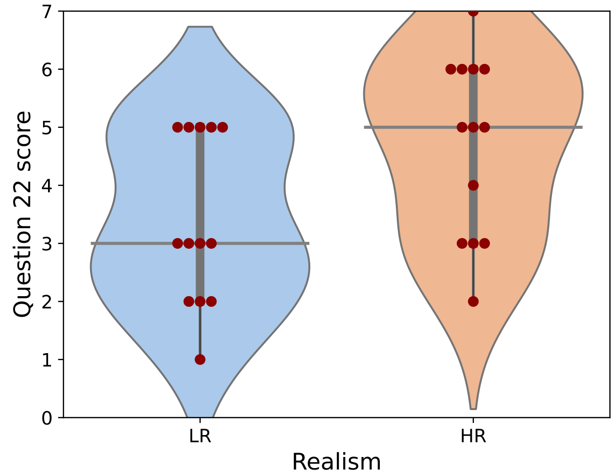

**Figure 8: Distribution of the overlap measured between the frames where an item was targeted with both HBF and GBF.**

the result is statistically significant ($p = 0.012$), while for GBF, although presenting data distributions similar to HBF, it could not be considered statistically significant ($p = 0.083$).

Figure 7 shows the scatter plot of the gaze-based focus (GBF) vs. head-based focus (HBF) measured with and without eye tracking.




(a)




(b)

**Figure 9: (a) Distribution of the total PQ questionnaire score for the LR and HR groups. The median is higher for the HR group as expected, but the difference is not statistically significant ($p = 0.5$ in a Mann-Whitney U test). (b) Distribution within the same groups of the single question "How much do you think the display's graphic quality interfered with or distracted you from completing the required tasks and activities?" on the Likert scale from 0 (not distracting) to 7 (very distracting). The median difference is statistically significant ($p < 0.05$ in a Mann-Whitney U test)**

The results show a positive correlation for both HR and LR groups, with Pearson coefficient $pcc = 0.47$ and $pcc = 0.68$ respectively and $pcc = 0.64$ considering both groups together. To measure the similarity of the two methods, we extracted from the data the number of frames in which a special object was targeted by either the GBF or the HBF. We calculated the overlap between the two as:

$$Overlap = \frac{frames_{HBF} \cap frames_{GBF}}{frames_{HBF} \cup frames_{GBF}}$$

where $frames_{HBF}$ is the set of frames where an object was targeted using the head-based method, $frames_{GBF}$ is the set of frames where an object was targeted using the gaze-based method, and the result is reported in Figure 8.

The PQ scores, both per category and total, are reported in Table 1. A statistical analysis using the Man-Whitney U test concluded that differences between the HR and LR groups are not significantly different for all the categories and the total scores ($p > 0.05$). The total score distribution for the two groups is shown in Figure 9a. Further analysis performed on the single items of the questionnaire also showed no significant differences, except for a question concerning the graphic quality of the display. In particular, the question asked, on a scale from 0 to 7, to state how much the graphic quality of the display distracted the user from performing the task. The result, reported in Figure 9b, shows a significant difference in the answers ($p = 0.035$), indicating the HR condition ($median = 5$) as more distracting than the LR one ($median = 3$).

## 5 Discussion

Regarding the first research question RQ1, results show that the realism of the scene has an effect on the attention behavior, highlighting that the time spent looking at the stimuli is significantly higher in LR environments.

This can depend on several causes, as there are many aspects defining graphical realism [10] that we have varied in the two settings. The outcomes of our experiments, however, reveal that the difference is not linked to the different levels of presence felt in the two environments. In fact, while there is a slightly higher average presence in the HR setting, as expected, there are no significantly different levels of presence measured with the PQ questionnaire. This contrasts with the finding of [16], who find statistically significant differences in presence in similar LR/HR settings (even if with non-highly significant). They conclude that this also impacts affective responses and perceptions. The fact that we do not measure a similar difference may be due to differences in the realization of the corresponding environments. We note that in [16] the authors use different models and scenes, even if of the same type, with fewer elements in the LR setting, while we focus on the realism of the illumination. In any case, while it may be true that some affective responses can be improved by higher realism, our results seem to suggest the opposite effect for attentive tasks. Moreover, this is interestingly confirmed by the answers to question 22 of PQ "How much do you think the display's graphic quality interfered with or distracted you from completing the required tasks and activities?". This is the only question where we recorded a significantly higher score in the HR group, meaning two important things: (i) the participants perceived clearly the difference in the graphics quality; (ii) the participants considered this "distracting" from the task activities. This seems to explain the reason for the significantly lower attention toward special objects in HR scenes.

In response to RQ2, the results indicate that the two different methods seem to capture in a similar way the differences across groups. They are obviously different, even if strongly correlated, as expected considering the literature [15]. While the temporal superimposition of the attention measured with the two methods is probably lower than the values expected from the analyses presented in [15], it should be considered that it is difficult to consider a real ground truth the value derived from gaze-based ray tracing as the angular section corresponding to the attention is not infinitely small and the measured gaze can be affected by different types of errors. The consistency of the results obtained with the two methods, in any case, suggests that head tracking can be used in VR studies to evaluate attentive saliency without the need for additional eye tracking, often unavailable in low-end headsets. Furthermore, head orientation can be tracked with external cameras [23] and this could make it possible to perform similar studies where the attention on special objects is also compared across real and virtual spaces.

## 6 Conclusion and future works

In this work, we designed an experiment to investigate the effect of graphic realism of IVEs in a context where VR is used to study cognitive behaviors such as attention. Compared to a reasonable level of realism, we found that a more stylized type of rendering affects users' attention and magnifies its effect in some cases. We consider this a potentially useful insight for future research using VR to measure human attention, also considering that a positive side-effect of designing and creating more stylized virtual scenarios usually results in a shorter development time and decreased costs. A limitation of the study is that it is clearly not possible to generalize the result for different types of behaviors or other styles of non-realistic rendering. This can definitely be a goal for future works. Regarding the technology implemented to track users' attention, we also demonstrated that it is possible to conduct such experiments using low-cost devices without integrated eye tracking. In fact, a head-tracking-based solution provided similar results in measuring attention. In future works, we plan to investigate different styles of rendering to provide more details on the specific aspects of the graphic that affect cognitive behaviors, as well as try different implementations of the methods to track the users' interest inside an IVE and compare the results with the findings of this work.

## Acknowledgments

This study was partially carried out within the PNRR research activities of the consortium iNEST (Interconnected North-Est Innovation Ecosystem) funded by the European Union Next-GenerationEU (Piano Nazionale di Ripresa e Resilienza (PNRR) – Missione 4 Componente 2, Investimento 1.5 – D.D. 1058 23/06/2022, ECS00000043).

## References

[1] Isayas Berhe Adhanom, Paul MacNeilage, and Eelke Folmer. 2023. Eye tracking in virtual reality: a broad review of applications and challenges. *Virtual Reality* 27, 2 (2023), 1481–1505.

[2] Robi Barranco Merino, Juan Luis Higuera-Trujillo, and Carmen Llinares Millán. 2023. The use of sense of presence in studies on human behavior in virtual environments: a systematic review. *Applied Sciences* 13, 24 (2023), 13095.

[3] Giulia Benvegnù, Alessandro Piva, Camilla Cadorin, Vanessa Mannari, Matteo Girondini, Angela Federico, Stefano Tamburin, and Cristiano Chiamulera. 2024. The effects of virtual reality environmental enrichments on craving to food in healthy volunteers. *Psychopharmacology* 241, 1 (2024), 49–60.

[4] Kai Biermeier, Ingrid Scharlau, and Enes Yigitbas. 2024. Measuring visual attention capacity across xreality. In *Proceedings of the 17th International Conference on PErvasive Technologies Related to Assistive Environments*. 98–106.

[5] Wanyu Chen and Haining Wang. 2022. Combining virtual reality and Eye Tracking to recognize users' aesthetic preference for product modeling. In *International Conference on Human-Computer Interaction*. Springer, 173–181.

[6] Valve Corporation. 2025. SRanipal - VIVE Eye and Facial Tracking. https://developer.vive.com/us/support/sdk/category_howto/vive-eye-tracking.html.

[7] crosire. 2015. ReShade post-processing injector. https://reshade.me.

[8] Eric DeMarbre, Jay Henderson, and Robert J Teather. 2024. Investigating Presence Across Rendering Style and Ratio of Virtual to Real Content in Mixed Reality. In *Proceedings of the 2024 ACM Symposium on Spatial User Interaction* (Trier, Germany) *(SUI '24)*. Association for Computing Machinery, New York, NY, USA, Article 23, 8 pages. doi:10.1145/3677386.3682098

[9] Francesco Foroni, Giulio Pergola, Georgette Argiris, and Raffaella I. Rumiati. 2013. The FoodCast research image database (FRIDa). *Frontiers in Human Neuroscience* Volume 7 - 2013 (2013). doi:10.3389/fnhum.2013.00051

[10] Guilherme Gonçalves, Hugo Coelho, Pedro Monteiro, Miguel Melo, and Maximino Bessa. 2022. Systematic review of comparative studies of the impact of realism in immersive virtual experiences. *Comput. Surveys* 55, 6 (2022), 1–36.

[11] Guilherme Gonçalves, Miguel Melo, Pedro Monteiro, Hugo Coelho, and Maximino Bessa. 2023. The role of different light settings on the perception of realism in virtual replicas in immersive Virtual Reality. *Computers & Graphics* 117 (2023), 172–182.

[12] Nikita Mae Harris, Robert W Lindeman, Clara Shui Fern Bah, Daniel Gerhard, and Simon Hoermann. 2023. Eliciting real cravings with virtual food: Using immersive technologies to explore the effects of food stimuli in virtual reality. *Frontiers in psychology* 14 (2023), 956585.

[13] Sungchul Jung and Robert W Lindeman. 2021. Perspective: Does realism improve presence in vr? suggesting a model and metric for vr experience evaluation. *Frontiers in virtual reality* 2 (2021), 693327.

[14] Weichen Liu, Gianna Andrade, Jurgen Schulze, Neal Doran, Kelly E Courtney, et al. 2022. Using virtual reality to induce and assess objective correlates of nicotine craving: Paradigm development study. *JMIR serious games* 10, 1 (2022), e32243.

[15] Jose Llanes-Jurado, Javier Marín-Morales, Masoud Moghaddasi, Jaikishan Khatri, Jaime Guixeres, and Mariano Alcañiz. 2021. Comparing eye tracking and head tracking during a visual attention task in immersive virtual reality. In *Human-Computer Interaction. Interaction Techniques and Novel Applications: Thematic Area, HCI 2021, Held as Part of the 23rd HCI International Conference, HCII 2021, Virtual Event, July 24–29, 2021, Proceedings, Part II 23*. Springer, 32–43.

[16] Mark Newman, Birgitta Gatersleben, Kayleigh J Wyles, and Eleanor Ratcliffe. 2022. The use of virtual reality in environment experiences and the importance of realism. *Journal of environmental psychology* 79 (2022), 101733.

[17] Bettina Olk, Alina Dinu, David J Zielinski, and Regis Kopper. 2018. Measuring visual search and distraction in immersive virtual reality. *Royal Society open science* 5, 5 (2018), 172331.

[18] Earnest Robot. 2020. AutoHand Unity Plugin. https://assetstore.unity.com/packages/tools/game-toolkits/auto-hand-vr-interaction-165323.

[19] Mel Slater, Pankaj Khanna, Jesper Mortensen, and Insu Yu. 2009. Visual realism enhances realistic response in an immersive virtual environment. *IEEE computer graphics and applications* 29, 3 (2009), 76–84.

[20] Valve Software. 2014. SteamVR. https://partner.steamgames.com/doc/features/steamvr/info.

[21] Valve Software. 2015. OpenVR SDK. https://partner.steamgames.com/doc/features/steamvr/openvr.

[22] Vinicius Souza, Anderson Maciel, Luciana Nedel, and Regis Kopper. 2021. Measuring presence in virtual environments: A survey. *ACM Computing Surveys (CSUR)* 54, 8 (2021), 1–37.

[23] Andrea Toaiari, Federico Cunico, Francesco Taioli, Ariel Caputo, Gloria Menegaz, Andrea Giachetti, Giovanni Maria Farinella, and Marco Cristani. 2023. Scene-pathy: capturing the visual selective attention of people towards scene elements. In *International Conference on Image Analysis and Processing*. Springer, 352–363.

[24] Tobii. 2025. XR Devzone. https://developer.tobii.com/xr/.

[25] Nikolaos Tsamitros, Stefan Gutwinski, Anne Beck, S Lange Mussons, Miriam Sebold, Robert Schöneck, Thomas Wolbers, Felix Bermpohl, Andreas Heinz, and Alva Lütt. 2024. Craving induction through virtual reality cue-exposure for patients with alcohol dependence in rehabilitation treatment. *Scientific reports* 14, 1 (2024), 1–10.

[26] Bob G. Witmer, Christian J. Jerome, and Michael J. Singer. 2005. The Factor Structure of the Presence Questionnaire. *Presence* 14, 3 (2005), 298–312. doi:10.1162/105474605323384654