Exploring how Saliency Affects Attention in Virtual Reality

Radiah Rivu[1], Ville Mäkelä[1,2,3], Mariam Hassib[1], Yomna Abdelrahman[1], and Florian Alt[1]

[1] Universität der Bundeswehr München, Germany, sheikh.rivu@unibw.de, mariam.hassib@unibw.de, yomna.abdelrahman@unibw.de, florian.alt@unibw.de
[2] LMU Munich, Munich, Germany, ville.maekelae@ifi.lmu.de
[3] University of Waterloo, Waterloo, Canada

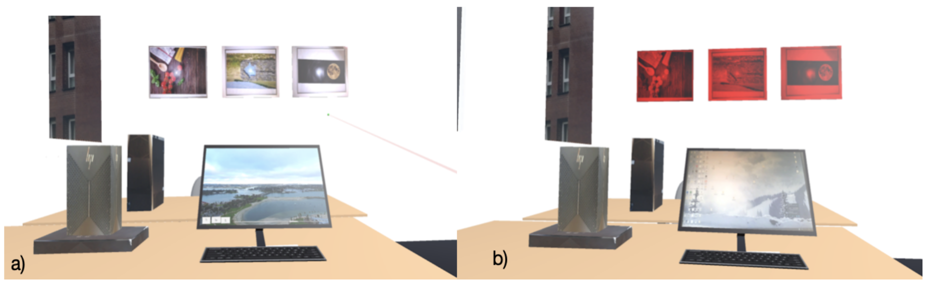

Fig. 1. We explored how different levels of saliency in a virtual environment affect the user's attention while performing tasks. a) Saliency condition 1: original paintings, b) Saliency condition 2: color of the paintings changed.

Abstract. We investigate how changes in the saliency of a Virtual Environment (VE) affect our visual attention during different tasks. In particular, we investigate if users are attracted to the most salient regions in the VE. This knowledge will help researchers design optimal VR environments, purposefully direct the attention of users, and avoid unintentional distractions. We conducted a user study (N=30) where participants performed tasks (video watching, object stacking, visual search, waiting) with two different saliency conditions in the virtual environment. Our findings suggest that while participants notice the differences in saliency, their visual attention is not diverted towards the salient regions when they are performing tasks.

Keywords: Virtual Reality, Saliency, Visual Attention

1 Introduction

Virtual Reality (VR) has become an important research topic in human-computer interaction. Over the years, researchers have studied various topics in VR like perception (immersion, presence, cognition) [7], novel input devices [15], locomotion [4], navigation [29], and avatars and virtual humans [23]. More recently, researchers have looked into how VR could be used to simulate real-world situations, for example, for research purposes [16, 17] and training [9], and looked in more detail into VR study methodologies [21].

However, our understanding is still limited as to how the design of the virtual environment (VE) influences human *visual attention* and behavior. Though factors affecting visual attention in the real world have been studied [14, 24], we know very little about its effects in VR. One key factor affecting visual attention is *saliency*, "the tendency of humans to be drawn to areas or objects that stand out amongst the background, when viewing natural scenes"[12].

Saliency is potentially a confounding variable in VR studies, and understanding its effects will help researchers design their experiment. Saliency is both a challenge and an opportunity for almost any VR experience, ranging from games [18] to cinematic VR [22]. Salient parts of the environment may distract players and viewers unintentionally. However, saliency can also be used to subconsciously direct users' attention towards relevant objects in the environment [3].

Prior work explores saliency in virtual reality [1, 6, 19, 25], but we lack an investigation into how saliency affects visual attention during various tasks, and whether this has an effect on task performance and behavior. Therefore, in this paper we conduct a preliminary exploration to understand if the saliency of VEs influences the user's visual attention while performing different tasks. Saliency is an umbrella term that consists of many factors such as color, shape, illumination and texture. Investigating each factor is out of scope for this paper, thus as a first-step towards research in this direction we study saliency in VR using *color*.

To this end, we conducted an experiment (N=30) where users performed tasks in VR (stacking 3D objects, searching for a specified object, watching a video, and waiting), while we manipulated the saliency of certain objects (paintings) in the environment. Our primary research question is: *Does saliency in a virtual environment affect visual attention and task performance?*

Our findings show that although participants notice changes in saliency, their visual attention is not diverted towards the salient regions when they are performing tasks. Participants also completed their tasks as efficiently in both saliency conditions, showing that saliency did not affect task performance.

2 Background

Saliency affects our visual attention [31]. Researchers have widely explored the effects of saliency in real-world environments and built saliency-based models. A popular saliency-based search model was presented by Itti and Koch [10]. The authors applied the model to a demanding search task and results show that saliency has a great impact on our visual attention and behaviour. Following this work, Underwood et al. [26] studied if saliency dominantly determines the visual attention given to objects, and if a cognitively demanding task can overtake this effect. Their findings suggest that when building saliency models, one should consider cognitive load as a metric [26].

Studies investigated whether saliency can be used to direct the user's attention. For example, Vig et al. [28] used machine learning to model how modifying the saliency of videos can redirect attention. Veas et al. [27] used visual saliency modulation to make subtle changes in videos, and were able to unobtrusively direct the viewer's attention to the modified regions. Meur et al. [13] found that saliency-based distortion of videos had only a moderate effect on which regions received the viewers' attention. Another application was to use saliency as a means to increase the security of gaze-based authentication mechanisms [2, 5].

In this paper, we are motivated to understand how saliency functions in VR. Due to the diversity in graphics, art style, lighting, and hardware, VR can be salient in many ways. We argue that exploring such effects in VR is important because we are yet unaware how this may affect user behaviour in VR.

3 Study: Investigation of Saliency in VR

To answer our research question, we conducted a user study in VR where participants performed tasks with a varying degree of saliency in the virtual environment. We implemented a virtual office room with paintings, tables, chairs and a screen. The VR environment was built using Unity and SteamVR. As the head-mounted display (HMD), we used the HTC VIVE Pro. The VIVE Pro has an integrated eye tracker from which we obtained gaze data.

3.1 Design

We used a within-subjects design where all participants experienced all conditions. We had two independent variables: 1) task and 2) level of saliency.

Tasks We selected the following four tasks (Figure 2), as these represent typical tasks in VR [20] and were also different in nature, ranging from physical tasks to more passive tasks. Participants performed each task twice (with and without change of saliency) in a counterbalanced order.

- **Manipulation Task:** Participants used the VR controllers to stack numbered cubes on the table (Figure 2a).
- **Search Task:** Participants searched for a hidden object (a red cube) in the room. They needed to use the controller to grab it and move it on the table (Figure 2b).
- **Video Watching Task:** Participants watched a one-minute video on the screen in front of them (Figure 2c).
- **Waiting Task:** This was a deceptive task where the participants were told to wait for the experimenter's instructions. This one-minute task was used to allow the participants to observe the environment (Figure 2d).

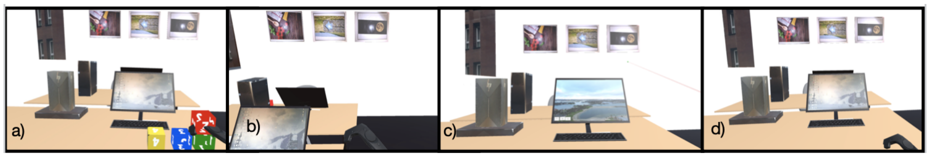

Fig. 2. Participants performed four kinds of tasks: a) Manipulation task where participants stack the numbered colored blocks, b) Visual search task where participants look for a hidden red cube, c) Video watching where participants watch a one minute video on the screen, d) Waiting task where participants wait for further instructions.

Saliency We used two saliency levels *(original, changed saliency)* as shown in Figure 1. Prior work explored several approaches to measure and quantify the saliency of an image or video [8]. To identify the salient regions of the virtual environment for our study, we used Saliency Toolbox [30]. The saliency maps show that when the color of the paintings is changed to red, they become more salient. We measured the saliency for three colors (red, blue and green) but red showed the highest increase, thus we selected it for our experiment.

3.2 Participants and Procedure

We recruited 30 participants (18 males, 12 females) with an average age of 27 (SD = 4.04) through university mailing lists and social media. 25 participants were students, five were professionals in different fields. 12 participants had no experience with VR or eye tracking.

Upon arrival, participants filled their demographic information (age, gender, background, experience with VR), signed a consent form, and were explained the tasks in the study. The task order was balanced using Latin square. After the study, they filled in the final questionnaire in which we asked open-ended questions about the objects they noticed in the environment, if they felt distracted while performing the tasks, and if their visual attention was diverted. The study lasted approximately 30 minutes. During the experiment, we recorded gaze data from the participants. Each participant was awarded 5 EUR.

4 Results

4.1 Attention towards the Salient Regions

We quantified and visualized the number of times participants glanced at the salient regions (the paintings) between the conditions. This was calculated using the number of times the gaze ray intersected with the observed virtual objects. To eliminate false positives, we only considered measurements from the moment the user's gaze left the initial starting point, similar to prior work [11]. After the study, we asked participants which objects in the room attracted their attention. 15 participants mentioned the paintings on the wall.

There was no significant difference in glance counts among any of the tasks between saliency levels (Table 1). This indicates that the saliency conditions of the VE did not affect the participants when they were focused on a task. We visualized the glances as colored dots as shown in Figures 3, 4, 5, and 6. The figures show that in both environments, users were focused on the tasks. We see greater variation for the *search task* and *waiting task*. This is because in *search task* participants were looking for a specified hidden object, and in the *waiting task* participants did not have anything particular to focus on.

Rating	Manipulation	Searching	Video	Waiting
Chi-Square $\chi^2(2)$	2.778	3.333	.043	2.133
ρ	.96	.068	.835	.144

Table 1. Result of the Friedman Test for gaze on the paintings, between two different saliency conditions.

Fig. 3. Gaze distribution for manipulation task: original (left) / changed (right) saliency.

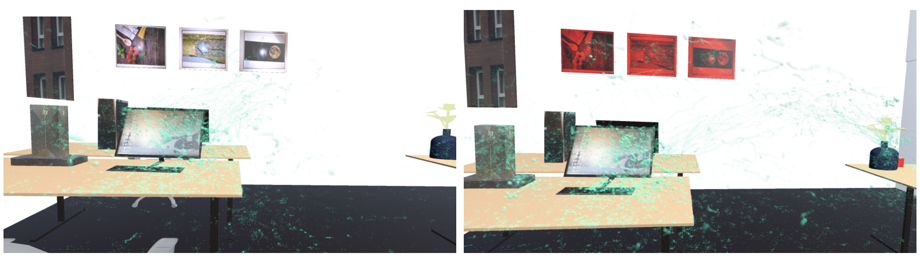

Fig. 4. Gaze distribution for search task: original (left) / changed (right) saliency.

4.2 Task Completion Time

The average task completion times show no significant difference between both saliency levels. Table 2 shows a comparison of the original and changed saliency and its effect on task completion time. Given the *waiting* and *video watching* task were of fixed length, we see no difference in completion time between both conditions (original and changed saliency). For the *manipulation* and *search task*, there is also no statistically significant difference between saliency levels (Table 2), despite a slight difference in the average completion times.

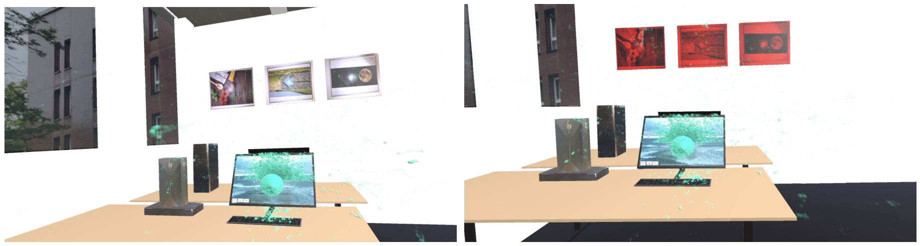

Fig. 5. Gaze distribution for video task: original (left) / changed (right) saliency.

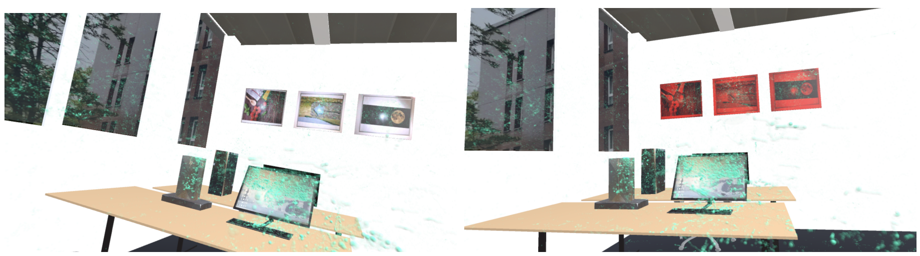

Fig. 6. Gaze distribution for waiting task: original (left) / changed (right) saliency.

Rating	Manipulation	Searching	Video	Waiting
Chi-Square $\chi^2(2)$	1.2000	.133	1.33	2.133
ρ	.273	.715	.715	.144

Table 2. Friedman Test between original and changed saliency condition for task completion time.

We also asked participants if they noticed anything unusual in the room. 16 out of 30 participants replied that they noticed the paintings on the wall changed color between tasks.

5 Discussion & Future Work

In our study, participants noticed the most salient VR regions in their field of view. This is in line with existing literature, that suggest that humans attend to the most salient regions of their surrounding [31]. Therefore, this preliminary investigation suggests that participants pay attention to saliency in virtual environments similar to real environments. We investigated how the salient regions would attract attention during different tasks. Tasks are frequently performed in VR (for example, user studies in VR, gaming, and other virtual experiences). Hence, it is important to understand how the virtual environment affects our behavior during VR experiences.

We found that the participants were unaffected by the saliency during the tasks. In both saliency conditions, participants did not noticeably glance at the salient regions during their tasks. They also performed their tasks equally fast in both settings. The short duration required to complete each task may have impacted the viewing behaviour of the participants. It is possible that if each task required a substantially longer time to complete then the participants would have been likely to also glance more at their surroundings during the tasks. In the future, we wish to investigate how different categories of tasks as well as task duration may be affected by saliency.

Saliency can be used to highlight information to users in VR, which is beneficial for many applications like virtual advertisements, storytelling, gaming, and shopping. However, based on our results, designers should be aware that salient regions do not gain the user's constant attention while they are focused on other tasks, even if the salient regions are located within their field of view. Therefore, designers should consider utilizing the more relaxed periods in their VR experiences (e.g., after a task has been finished) to guide users through saliency.

Our early investigation had some limitations and therefore there is room for more in-depth studies in the future. First, we only explored one type of subtle change in saliency and we only focused on the most salient objects in the environment (the paintings on the wall). The saliency of a VE can be manipulated in various other ways — and to a varying extent — and these should be investigated in the future. In particular, research should investigate the effect of combining other forms of saliency to understand its effect in VR, mainly combining color with shape and illumination as all these play a role in the design of VR environments. Second, visual attention could be further investigated by considering additional measures such as eye blinking and fixation time. Third, we wish to explore visual attention during a more diverse set of tasks. In particular, we plan to include VR tasks of varying cognitive requirements to see if saliency plays a bigger role in cognitively demanding tasks.

6 Conclusion

To explore the effect of saliency on visual attention and task performance in VR, we conducted a preliminary study where participants performed several tasks in VR while we manipulated the saliency of the environment. We learned that although participants do notice different saliency levels in the VE, it has no significant effect on visual attention or task performance. Rather, participants focus on the objects relevant to their task, and are more prone to detecting the salient regions in the environment between the tasks. We believe that our work helps us understand how saliency affects attention in VR. For example, designers of virtual environments and virtual experiences who want to use saliency changes to guide users, could consider leveraging the downtime between tasks for more effective guidance.

Acknowledgements. We thank the Rec Room tutor who aided in recruiting participants for our study. The presented work was funded by the German Research Foundation (DFG) under project no. 316457582 and 425869382.

References

1. D. Albayrak, M. B. Askin, T. K. Capin, and U. Celikcan. 2019. Visual Saliency Prediction in Dynamic Virtual Reality Environments Experienced with Head-Mounted Displays: An Exploratory Study. In *2019 International Conference on Cyberworlds (CW)*. 61–68.
2. Florian Alt, Mateusz Mikusz, Stefan Schneegass, and Andreas Bulling. 2016. Memorability of Cued-Recall Graphical Passwords with Saliency Masks. In *Proceedings of the 15th International Conference on Mobile and Ubiquitous Multimedia (MUM '16)*. Association for Computing Machinery, New York, NY, USA, 191–200. DOI: http://dx.doi.org/10.1145/3012709.3012730
3. Reynold Bailey, Ann McNamara, Nisha Sudarsanam, and Cindy Grimm. 2009. Subtle Gaze Direction. *ACM Trans. Graph.* 28, 4, Article 100 (Sept. 2009), 14 pages. DOI:http://dx.doi.org/10.1145/1559755.1559757
4. Costas Boletsis. 2017. The new era of virtual reality locomotion: A systematic literature review of techniques and a proposed typology. *Multimodal Technologies and Interaction* 1, 4 (2017), 24.
5. Andreas Bulling, Florian Alt, and Albrecht Schmidt. 2012. Increasing the Security of Gaze-Based Cued-Recall Graphical Passwords Using Saliency Masks. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '12)*. Association for Computing Machinery, New York, NY, USA, 3011–3020. DOI:http://dx.doi.org/10.1145/2207676.2208712
6. A. De Abreu, C. Ozcinar, and A. Smolic. 2017. Look around you: Saliency maps for omnidirectional images in VR applications. In *2017 Ninth International Conference on Quality of Multimedia Experience (QoMEX)*. 1–6.
7. Julia Diemer, Georg W Alpers, Henrik M Peperkorn, Youssef Shiban, and Andreas Mühlberger. 2015. The impact of perception and presence on emotional reactions: a review of research in virtual reality. *Frontiers in psychology* 6 (2015), 26.
8. Y. Fang, J. Wang, M. Narwaria, P. Le Callet, and W. Lin. 2014. Saliency Detection for Stereoscopic Images. *IEEE Transactions on Image Processing* 23, 6 (2014), 2625–2636.
9. Dominic Gorecky, Mohamed Khamis, and Katharina Mura. 2017. Introduction and establishment of virtual training in the factory of the future. *International Journal of Computer Integrated Manufacturing* 30, 1 (2017), 182–190. DOI: http://dx.doi.org/10.1080/0951192X.2015.1067918
10. Laurent Itti and Christof Koch. 2000. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision research* 40, 10-12 (2000), 1489–1506.
11. T. Judd, K. Ehinger, F. Durand, and A. Torralba. 2009. Learning to predict where humans look. In *2009 IEEE 12th International Conference on Computer Vision*. 2106–2113.
12. Kathryn Koehler, Fei Guo, Sheng Zhang, and Miguel P Eckstein. 2014. What do saliency models predict? *Journal of vision* 14, 3 (2014), 14–14.
13. Olivier Le Meur, Alexandre Ninassi, Patrick Le Callet, and Dominique Barba. 2010. Overt visual attention for free-viewing and quality assessment tasks: Impact of the regions of interest on a video quality metric. *Signal Processing: Image Communication* 25, 7 (2010), 547–558.
14. Gordon D Logan and Robert D Gordon. 2001. Executive control of visual attention in dual-task situations. *Psychological review* 108, 2 (2001), 393.
15. Christoph Maggioni. 1993. A novel gestural input device for virtual reality. In *Proceedings of IEEE Virtual Reality Annual International Symposium*. IEEE, 118–124.

16. Ville Mäkelä, Rivu Radiah, Saleh Alsherif, Mohamed Khamis, Chong Xiao, Lisa Borchert, Albrecht Schmidt, and Florian Alt. 2020. Virtual Field Studies: Conducting Studies on Public Displays in Virtual Reality. In *Proceedings of the 2020 CHI Conference on Human Factors in Computing Systems (CHI '20)*. Association for Computing Machinery, New York, NY, USA, 1–15. DOI: http://dx.doi.org/10.1145/3313831.3376796
17. Mehdi Moussaïd, Mubbasir Kapadia, Tyler Thrash, Robert W Sumner, Markus Gross, Dirk Helbing, and Christoph Hölscher. 2016. Crowd behaviour during high-stress evacuations in an immersive virtual environment. *Journal of The Royal Society Interface* 13, 122 (2016), 20160414.
18. Néstor Ordaz, David Romero, Dominic Gorecky, and Héctor R Siller. 2015. Serious games and virtual simulator for automotive manufacturing education & training. *Procedia Computer Science* 75 (2015), 267–274.
19. Oyewole Oyekoya, William Steptoe, and Anthony Steed. 2009. A saliency-based method of simulating visual attention in virtual scenes. In *Proceedings of the 16th ACM Symposium on Virtual Reality Software and Technology*. 199–206.
20. JW Regian, Wayne L Shebilske, and John M Monk. 1992. Virtual reality: An instructional medium for visual-spatial tasks. *Journal of Communication* (1992).
21. Radiah Rivu, Ville Mäkelä, Sarah Prange, Sarah Delgado Rodriguez, Robin Piening, Yumeng Zhou, Kay Köhle, Ken Pfeuffer, Yomna Abdelrahman, Matthias Hoppe, Albrecht Schmidt, and Florian Alt. 2021. Remote VR Studies - A Framework for Running Virtual Reality Studies Remotely Via Participant-Owned HMDs. *CoRR* abs/2102.11207 (2021). https://arxiv.org/abs/2102.11207
22. Sylvia Rothe and Heinrich Hußmann. 2018. Guiding the Viewer in Cinematic Virtual Reality by Diegetic Cues. In *Augmented Reality, Virtual Reality, and Computer Graphics*, Lucio Tommaso De Paolis and Patrick Bourdot (Eds.). Springer International Publishing, Cham, 101–117.
23. Anthony Steed, Ye Pan, Fiona Zisch, and William Steptoe. 2016. The impact of a self-avatar on cognitive load in immersive virtual reality. In *2016 IEEE Virtual Reality (VR)*. IEEE, 67–76.
24. Yaoru Sun and Robert Fisher. 2003. Object-based visual attention for computer vision. *Artificial intelligence* 146, 1 (2003), 77–123.
25. M. Sung and S. Choi. 2017. Selective Anti-Aliasing for Virtual Reality Based on Saliency Map. In *2017 International Symposium on Ubiquitous Virtual Reality (ISUVR)*. 16–19.
26. Geoffrey Underwood, Tom Foulsham, Editha van Loon, and Jean Underwood. 2005. Visual attention, visual saliency, and eye movements during the inspection of natural scenes. In *International Work-Conference on the Interplay Between Natural and Artificial Computation*. Springer, 459–468.
27. Eduardo E. Veas, Erick Mendez, Steven K. Feiner, and Dieter Schmalstieg. 2011. Directing Attention and Influencing Memory with Visual Saliency Modulation. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '11)*. Association for Computing Machinery, New York, NY, USA, 1471–1480. DOI:http://dx.doi.org/10.1145/1978942.1979158
28. Eleonora Vig, Michael Dorr, and Erhardt Barth. 2011. Learned saliency transformations for gaze guidance. In *Human Vision and Electronic Imaging XVI*, Vol. 7865. International Society for Optics and Photonics, 78650W.
29. Norman G Vinson. 1999. Design guidelines for landmarks to support navigation in virtual environments. In *Proceedings of the SIGCHI conference on Human Factors in Computing Systems*. 278–285.

30. Dirk Walther and Christof Koch. 2006. Modeling attention to salient proto-objects. *Neural networks* 19, 9 (2006), 1395–1407.
31. FengJun Zhang, GuoZhong Dai, and Xiaolan PENG. 2016. A survey on human-computer interaction in virtual reality. *Scientia Sinica Informationis* 46, 12 (2016), 1711–1736.