Exploring Gaze Behavior in Immersive VR under Cognitive Load

Jaime Bielsa Jorge Pina Ana Serrano Daniel Martin*

Universidad de Zaragoza - I3A, Spain

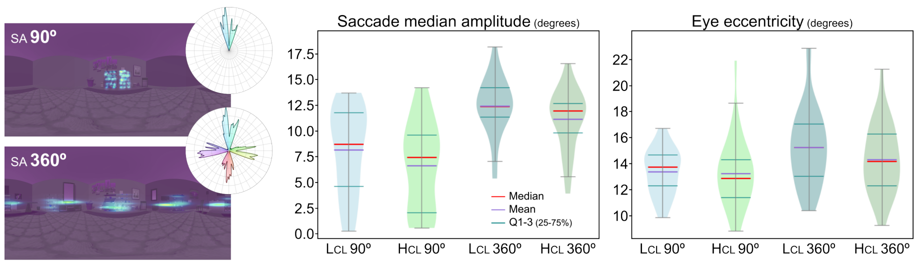

Figure 1: **Left:** Saliency maps for the visual search experiment from Pina et al. [5], for the 90° and 360° search area, with their corresponding radial plots as insets. Both correspond to conditions of low cognitive load. **Right:** Saccade median amplitude and eye eccentricity, averaged for all users per experimental condition (i.e., low and high cognitive load for 90° and 360° search areas). Our analyses suggest a change in gaze when exploring larger regions and when a higher cognitive level is present (see Table 1).

Abstract

Understanding human visual attention is a cornerstone for creating engaging experiences. This task remains far from trivial, especially within immersive environments where users perform multiple multisensory tasks under varying levels of cognitive load. In this work, we analyze gaze data from users performing a multisensory visual search experiment in VR, focusing on the differences in visual behavior across various cognitive load levels and search areas. Results suggest a gaze stabilization strategy translated into significantly smaller saccades and eye eccentricity, which may relate to sensorimotor economy.

Index Terms: Virtual Reality, Eye-tracking, Gaze, Saliency

1 Introduction

Cognitive load (CL) refers to the mental resources allocated to perform a task, a concept rooted in Sweller's foundational work [7]. Classical literature has investigated how CL influences behavior by manipulating task difficulty or adding secondary tasks, and by measuring its effects in physiological responses or reaction times. However, these studies have predominantly relied on traditional, desktop-based setups, which do not reflect the complex demands of fully immersive environments.

Virtual reality (VR) offers a significantly different environment: Wide fields of view with context in the 360° around the user and complete user control of the camera create conditions where visual attention becomes both less restricted and more representative of real-world situations. In these immersive environments, users cannot perceive the whole environment at a glance, and thus heavily rely on coordinated eye and head movements to guide attention. When multisensory tasks are introduced, these attentional demands may compete for limited cognitive resources.

In this work, we study how cognitive load, induced by multisensory inputs, and different search areas affect gaze behavior within immersive VR. We leverage the dataset collected during the experiment by Pina et al. [5], who captured several physiological markers, objective and subjective metrics, and eye-tracking data during a visual search task under two different cognitive load levels (low and high) and within different search areas (90° and 360°). We focus on such eye-tracking data, and aim to analyze whether differences in gaze behavior arise within different cognitive load levels or search areas, and whether those changes might relate to the performance results reported in the original work.

Our results show significant differences (i) in most of the analyzed gaze metrics within different search areas, probably due to the head rotations allowed in the 360° case, and (ii) in two metrics for different cognitive load levels (saccade amplitude and eye eccentricity), both of them being smaller for high cognitive load cases. These results suggest a gaze stabilization strategy, which is consistent with the sensorimotor economy that has been reported in attention research [4].

2 Dataset and Data Processing

In this work, we resort to the dataset collected by Pina et al. [5]. The data were collected during an experiment in which users were instructed to perform a visual search task during four different conditions, resulting from the combination of two factors: low or high cognitive load (LCL and HCL, respectively), and smaller (90°) or larger (360°) search areas (SA). High CL segments included additional auditory multitasking. Within the 90° search area condition, all items fit within the user's field of view. In contrast, in the 360° search area condition, objects were located across the entire room, and thus participants had to perform rotations. The entire experimental procedure was conducted using the Varjo-XR4 headset, and was completed by 35 participants (12 of them identified as female).

The dataset contains physiological markers (such as electrocardiogram or photoplethysmography, among others), objective performance metrics, subjective questionnaires, and eye-tracking data. In this work, we focus on the latter. The eye-tracking data were captured with the built-in eye-tracker from the headset, and include position and rotation of the headset, interpupillary distance, focus distance, and stability; and gaze direction, pupil and iris dilation, and eye openness for both eyes, together with an additional gaze direction that combines both eyes (i.e., Cyclopean eye). We take the latter and first compute the gaze UV coordinates for every frame by casting a ray (i.e., the gaze direction projected from the position and rotation of the headset) and then compute gaze velocity. With that, we compute a classification of eye-tracking data using a Velocity-Threshold Identification (I-VT) algorithm [1] (velocity threshold of 700°/sec. and thresholding of 75% of eye openness), and focus on fixations and saccades for our posterior analyses, as they have been reported to be good behavioral correlates [3, 8].

*e-mail: danims@unizar.es

3 Data Analysis

Following previous approaches [6], we first compute saliency maps for each experimental condition (i.e., each search area, SA; and each cognitive load level, CL) by convolving discrete fixation maps with a Gaussian kernel with $\sigma = 1^\circ$, the estimated size of the human fovea. Fig. 1 shows the maps corresponding to low CL conditions for both SA, together with radial plots also computed from gaze longitude. The saliency map for the 90° SA illustrates how attention is sharply focused on the items to be found during the experiment, while the saliency map for the 360° SA shows a blurrier distribution, mostly due to users being required to rotate around the virtual room to find objects constantly. To analyze whether such distributions vary for both cognitive load levels within the same SA, we resort to Pearson's correlation (CC, 1 for perfect and -1 for inverse correlations) and Kullback-Leibler divergence (KLD, which measures divergence between two distributions, i.e., the lower the better). Our results ($CC_{90^\circ} = 0.861 \pm 0.058$, $CC_{360^\circ} = 0.885 \pm 0.060$; $KLD_{90^\circ} = 0.420 \pm 0.240$, $KLD_{360^\circ} = 0.352 \pm 0.219$) suggest that there are no prominent variations in gaze distributions between low and high cognitive load levels within the same SA.

Since saliency distributions do not show significant differences, we focus on analyzing whether low-level gaze metrics differ based on the level of cognitive load. Inspired by Malpica et al. [3], we focus on dwell time (total time that a user is fixating during a condition), eye eccentricity (the average angular difference between the gaze point and the viewport center), head orientation entropy (Shannon entropy of the orientation of the head), the number and duration of fixations, and the number, duration, and amplitude (i.e., the angle the gaze describes between fixations) of saccades.

We first analyzed whether our data follows a normal distribution. A Shapiro-Wilk test shows that more than half of our metrics do not follow a normal distribution ($p \leq 0.05$). Since we are dealing with nonparametric data and multiple factors, we apply the Aligned Rank Transform (ART) for non-parametric factorial ANOVA [2, 9], which adds a preprocessing step after which common ANOVA procedures can be used. In our case, for every ANOVA model, each studied metric is the dependent variable, SA and CL are fixed factors, and users are included as random effects. Our results (Tab. 1) show that seven of the eight metrics are significantly different between search areas: For 360° settings, there are more gaze events. Fixations are shorter in duration (half the time in 360°), and saccades, while of equal duration, are wider. All of this highlights how participants had to scan larger areas and devoted less time focusing on each object until finding the required one.

The effect of cognitive load on gaze is less prominent. Our results show a significant effect on saccade amplitude and eye eccentricity. Both decrease during high CL segments with respect to low CL ones. The distribution of both metrics for each CL level and SA can be seen in the violin plots of Fig. 1. This suggests that the presence of an auditory secondary task does not directly affect the number of gaze events but rather the distance between them. A decrease in the saccade amplitude suggests that less space is scanned with each saccade, which aligns with the original work's report of fewer items found along the high CL runs [5]. We hypothesize that, since the auditory task is absorbing some cognitive resources, participants are not able to visually scan the scene so rapidly. The decrease in eye eccentricity aligns with this too, since higher cognitive demands reduce the resources available for visual exploration, ultimately resulting in narrower oculomotor behavior. Overall, these results seem to indicate that under high CL, participants appear to stabilize gaze by aligning eyes more closely with head direction and reducing saccade amplitude, a strategy consistent with the sensorimotor economy already known in attention research [4].

Metric	p-value		ES - Partial η^2	
	CL	SA	CL	SA
Dwell time	0.821	$<$ **0.001**	0.001	0.690
Number of fixations	0.398	$<$ **0.001**	0.007	0.650
Number of saccades	0.172	$<$ **0.001**	0.020	0.280
Fixation median duration	0.288	$<$ **0.001**	0.010	0.710
Saccade median duration	0.165	0.183	0.020	0.020
Saccade median amplitude	**0.001**	$<$ **0.001**	0.110	0.540
Eye eccentricity	**0.024**	$<$ **0.001**	0.050	0.250
Head orientation entropy	0.069	$<$ **0.001**	0.030	0.810

Table 1: Results from the Aligned Rank Transform (ART) for non-parametric factorial ANOVA for different gaze metrics for different search area (SA) and cognitive load (CL) levels. Significant differences are boldfaced, and effect sizes (partial η^2) are included for each of them. Search area strongly influences gaze behavior.

4 Conclusion and Future Work

In this work, we have analyzed gaze behavior within multisensory multitasking in VR. We have found a strong effect of the search area, as well as an apparent gaze stabilization strategy, which is visible in smaller saccade amplitudes and a smaller eye eccentricity. These results could partially explain the decrease in performance in the original experiment with higher CL levels. Many future avenues remain open: Studying more diverse gaze features (such as blinks or pupil dilation) could help better understand the impact of cognitive load and search area, and extending these relationships to the physiological signals from the dataset would also allow a better understanding of the correlation between gaze and other biological markers. Finally, utilizing this knowledge to inform real-time applications to adjust difficulty and increase performance in demanding scenarios remains an interesting avenue.

Acknowledgments

This research was funded by grant PID2022-141539NB-I00, funded by MICIU/AEI/10.13039/501100011033 and by ERDF, EU; by the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement nº 101220555); and by the Aragon Institute for Engineering Research (I3A) through the Impulso program. J. Pina was supported by an FPI predoctoral grant (PRE2023-UZ-16).

References

[1] E. Bernal-Berdun, D. Martin, S. Malpica, P. J. Perez, D. Gutierrez, B. Masia, and A. Serrano. D-sav360: A dataset of gaze scanpaths on 360° ambisonic videos. *IEEE TVCG*, 29(11):4350–4360, 2023. 1

[2] D. Kim, J. Kim, J.-E. Shin, B. Yoon, J. Lee, and W. Woo. Effects of virtual room size and objects on relative translation gain thresholds in redirected walking. In *IEEE VR*, pp. 379–388. IEEE, 2022. 2

[3] S. Malpica, D. Martin, A. Serrano, D. Gutierrez, and B. Masia. Task-dependent visual behavior in immersive environments: A comparative study of free exploration, memory and visual search. *IEEE TVCG*, 29, 2023. 1, 2

[4] S. Pannasch and B. M. Velichkovsky. Distractor effect and saccade amplitudes: Further evidence on different modes of processing in free exploration of visual images. *Visual Cognition*, 17(6-7), 2009. 1, 2

[5] J. Pina, E. Bernal-Berdun, D. Martin, S. Malpica, C. Real, A. Barquero, P. Armanac-Julian, J. Lazaro, A. Martin-Yebra, B. Masia, and A. Serrano. A Comprehensive Analysis of the Influence of Cognitive Load on Physiological Signals in Virtual Reality. In *ISMAR*, 2025. 1, 2

[6] V. Sitzmann, A. Serrano, A. Pavel, M. Agrawala, D. Gutierrez, B. Masia, and G. Wetzstein. Saliency in VR: How do people explore virtual environments? *IEEE TVCG*, 24(4):1633–1642, 2018. 2

[7] J. Sweller. Cognitive load during problem solving: Effects on learning. *Cognitive Science*, 12(2):257–285, 1988. 1

[8] D. Tovar, J. Wilmott, X. Wu, D. Martin, M. Proulx, D. Lindberg, Y. Zhao, O. Mercier, and P. Guan. Identifying behavioral correlates to visual discomfort. *ACM Trans. on Graphics (TOG)*, 43(6), 2024. 1

[9] J. O. Wobbrock, L. Findlater, D. Gergle, and J. J. Higgins. The aligned rank transform for nonparametric factorial analyses using only anova procedures. In *Proc. SIGCHI*, pp. 143–146, 2011. 2