Thinking on Your Feet: Enhancing Foveated Rendering in Virtual Reality During User Activity

DOI:
10.5220/0012459300003660

Document Version
Final published version

Thinking on Your Feet: Enhancing Foveated Rendering in Virtual Reality During User Activity

David Petrescu[1] [a], Paul A. Warren[2] [b], Zahra Montazeri[1] [c], Gabriel Strain[1] [d] and Steve Pettifer[1] [e]

[1] *University of Manchester, Human-Computer Systems, U.K.*

[2] *University of Manchester, Virtual Reality Research (VR2) Facility, Division of Psychology, Communication and Human Neuroscience, U.K.*

Keywords: Foveated Rendering, Variable Rate Shading, Psychophysics, Movement, Attention, Virtual Reality.

Abstract: As prices fall, VR technology is experiencing renewed levels of consumer interest. Despite wider access, VR still requires levels of computational ability and bandwidth that often cannot be achieved with consumer-grade equipment. Foveated rendering represents one of the most promising methods for the optimization of VR content while keeping the quality of the user's experience intact. The user's ability to explore and move through the environment with 6DOF separates VR from traditional display technologies. In this work, we explore if the type of movement (Active versus Implied) and attentional task type (Simple Fixations versus Fixation, Discrimination, and Counting) affect the extent to which a dynamic foveated rendering method using Variable Rate Shading (VRS) optimizes a VR scene. Using psychophysics methods we conduct user studies and recover the Maximum Tolerated Diameter (MTD) at which users fail to notice drops in quality. We find that during self-movement, performing a task that requires more attention masks severe shading reductions and that only 31.7% of the headset's FOV is required to be rendered at the native pixel sampling rate.

1 INTRODUCTION

VR is experiencing a significant increase in interest (Harley, 2020). With the rise of affordable and powerful chips that can generate realistic VR and Augmented Reality (AR) content, and Mixed Reality (MR) gaining popularity, photorealistic content becomes more desirable. However, rendering such environments is expensive, necessitating methods of reducing computational load and bandwidth usage. With resolutions and refresh rates of Head-Mounted Displays (HMDs) continuing to increase, rendering requirements become proportionally more expensive. Shading operations require significant computations performed on the Graphical Processing Unit (GPU). The Human Visual System (HVS) is particularly adept at resolving fine detail in the foveal region (i.e. focal point) but becomes less proficient as eccentricity increases; therefore, rendering techniques that exploit perceptual characteristics without affecting the overall experience present a solution to the increased computational load typical of VR applications.

During VR usage, users are commonly moving through a virtual environment either via active movement (i.e. under their own steam - locomotion) or passively via implied movement (e.g. simulated movement in a vehicle, using the controllers to navigate). Furthermore, users are often intentionally engaged in a task (e.g. finding enemies, searching for items). With respect to locomotion, there is evidence that the nature and complexity of neural processing differs for active versus passive movement. More specifically, the complex interplay between vestibular, proprioceptive, visual, and efference copy systems in active movement are required for the correct interpretation of a stable environment (Warren et al., 2022). With respect to task engagement, the deployment of what is referred to in psychological science as *overt attention* (see Section 2.3) is required. Based on previous work we know that:

- Retinal motion and/or motion from self-induced movement cause a decrease in visual sensitivity to fine detail (Murphy, 1978; Braun et al., 2017).
- Variable Rate Shading (VRS) artifacts in a fixed-foveated algorithm are less conspicuous during active movement compared to implied movement (Petrescu et al., 2023a).

[a] https://orcid.org/0000-0002-8949-7265
[b] https://orcid.org/0000-0002-4071-7650
[c] https://orcid.org/0000-0003-0398-3105
[d] https://orcid.org/0000-0002-4769-9221
[e] https://orcid.org/0000-0002-1809-5621

Petrescu, D., Warren, P., Montazeri, Z., Strain, G. and Pettifer, S.
Thinking on Your Feet: Enhancing Foveated Rendering in Virtual Reality During User Activity.
DOI: 10.5220/0012459300003660

In *Proceedings of the 19th International Joint Conference on Computer Vision, Imaging and Computer Graphics Theory and Applications (VISIGRAPP 2024) - Volume 1: GRAPP, HUCAPP and IVAPP*, pages 140-150
ISBN: 978-989-758-679-8; ISSN: 2184-4321

- Attention can be modulated by varying task difficulty, affecting the extent to which foveated graphics can be used (Krajancich et al., 2023).

The issues highlighted above suggest that there is potential for further improving foveated rendering algorithms by considering both user movement and task engagement. In particular, there may be greater scope for lower-quality rendering (and thus lower required levels of GPU computation) for users who are both moving and engaged in a task. In the present study, we investigate this potential. We achieve this by using readily available information about user movement and task type - a direct approach that does not rely on models of vision.

2 BACKGROUND

2.1 Foveated Rendering

Foveated graphics are a family of optimization methods that rely on uneven retinal sensitivity to detail across the visual field. Due to the physical distribution of photoreceptors on the retina (Curcio and Allen, 1990), sensitivity to fine detail decreases as a function of eccentricity from the focal point; foveated graphics exploit this property by reducing rendering quality as a function of retinal acuity falloff. Psychophysical studies have confirmed that these retinal properties result in lower visual acuity and spatio-temporal contrast sensitivity in the visual periphery (Krajancich et al., 2021; Mantiuk et al., 2022). Foveated graphics can be *gaze-contingent* (eye-tracked - degradation follows gaze) or *fixed* to the center of the display. For a state-of-the-art report refer to Wang et al. (2023).

Guenter et al. (2012) approximated the retinal falloff to a linear function in Minimum Angle of Resolution (MAR) units (acuity reciprocal). They showed that by using foveated rendering with an aggressive degradation MAR slope, sampling performance could be increased by a factor of five. Foveated rendering is especially attractive for VR HMDs, especially with eye-tracking technology becoming more readily available, and has been successfully implemented in previous work (Patney et al., 2016; Vaidyanathan et al., 2014). Mantiuk et al. (2021) introduced a state-of-the-art visual difference metric based on a comprehensive visual model that accounts for eccentricity and sensitivity to spatio-temporal artifacts, which can be used to assess foveated methods or other degradation techniques. Models of vision have since been used to enhance foveated graphics whilst accounting for contrast sensitivity (Tursun and Didyk, 2022; Tursun et al., 2019). Guiding samples in ray-tracing algorithms can also benefit from foveated graphics, with this being implemented successfully by Weier et al. (2016). Other applications of foveated graphics include increasing the perceived dynamic range of the image (E. Jacobs et al., 2015), achieving more efficient power consumption by manipulating how eccentricity affects color perception (Duinkharjav et al., 2022), and guiding level-of-detail (LOD) strategies (Luebke and Hallen, 2001).

2.2 Variable Rate Shading

VRS (NVIDIA, 2018) is a technology supported by the Nvidia Turing architecture that enables variable control of shading across different parts of an image independently; this technique confers more granularity than conventional shading techniques. VRS has many similarities to coarse pixel shading – a method used previously for shading simplifications and introduced as part of a foveated rendering method in Vaidyanathan et al. (2014). The rendered images are divided into 16×16px tiles which can be shaded in a multitude of configurations. The rasterization process is not altered when VRS is activated, ensuring that no additional aliasing artifacts are introduced and sharp edges are preserved. The shading rate of the 16×16px tile can have a uniform configuration: $\{2\times2, 4\times4\}$ or a non-uniform one: $\{1\times2, 2\times1, 2\times4, 4\times2\}$. In order to minimize calls to the fragment shader, the final pixel appearance is derived from only one value based on the configurations. It is important to mention that this downsampling results in visual artifacts that can be scene-dependent, with certain textures (e.g. checkered patterns) being more affected by VRS degradations. In our experiment, we use VRS4×4 for maximum optimization benefits.

2.3 Attention

Recently, considerable research effort in the perceptual graphics community has shifted from low-level visual processing (e.g. contrast sensitivity) to higher-order processes such as visual attention. Attention is often characterized as a 'spotlight' for the focusing of cognitive resources (Cave and Bichot, 1999). We, however, aim to study dynamic environments; there is evidence that, when scene movement is present, attention is structured in terms of perceptual groups as opposed to predefined areas (Driver and Baylis, 1989).

The term *overt attention* typically refers to the process of directly orienting one's gaze to objects of interest, whereas *covert attention* refers to the mechanisms of directing attention to a part (or potentially multiple parts) of the environment without moving the eyes. Covert attention is thought to precede overt attention and guide subsequent eye movements/fixations. In our experiment, we are interested in endogenous overt attention, specifically the ability to voluntarily monitor information that appears at a given location and with the participant performing a certain task.

Attention has been previously studied in the context of perception and visual sensitivity. Whilst there is evidence that attending overtly to a focal point improves performance and acuity in the foveal region (Kandel et al., 2012), it has also been recently shown that the presence of a task that requires increased mental load significantly reduces one's visual acuity (Mahjoob et al., 2022). Moreover, the deployment of attention is also thought to reduce contrast sensitivity (Huang and Dobkins, 2005). Recently, Krajancich et al. (2023) created an attention-aware model of contrast sensitivity. By running extensive user studies, they provided strong evidence that the additional presence of a task which requires an increased level of attentional load further lowers the ability to resolve fine detail in the periphery (i.e. the degradation area can be increased when attention is deployed).

2.4 Movement and Optimization

Denes et al. (2020) created a vision model accounting for resolution and refresh rate that improved the visual quality of predictable and unpredictable motion. However, they only studied on-screen motion and did not consider the source of movement. VRS has also been used to great effect alongside models of motion and visual masking in order to reduce rendering load and account for motion artifacts (Jindal et al., 2021; Yang et al., 2019). More recently, neural network super-resolution techniques accounting for temporal changes were used to great effect to optimize foveated graphics in VR (Ye et al., 2023).

Suchow and Alvarez (2011a,b) explored a phenomenon termed 'silencing' which highlights that the presence of motion in the background impairs the ability to discern between qualitative properties of the foreground stimuli (luminance, color, size and shape).

Regarding the effects of type of movement on degradation artifacts, Petrescu et al. (2023b) showed that users' head rotations can be used to drive a LOD simplification algorithm. Ellis and Chalmers (2006) used dynamic scaling of the foveated region (at multiple sampling rates) scaled to the vestibular response produced by translational movement using a 6-Degrees of Freedom (DOF) motion pod and found significant scope for optimization. Additionally, Petrescu et al. (2023a) examined the effects of VRS-based fixed-foveated rendering degradations during instances of self-movement and implied movement, finding that self-movement reduced sensitivity to more pronounced foveated rendering settings. Moreover, Lisboa et al. (2023) recently showed that rectangle-mapped foveated rendering can be enhanced during instances of implied user movement. Using depictions of indoor and outdoor environments, they showed that when the user is being moved through the VR environment at higher velocities, the severity of the foveated rendering algorithm can be significantly increased compared to a stationary user without affecting the visual experience.

In the current study, we build upon the research presented in Krajancich et al. (2023), Petrescu et al. (2023a), and Ellis and Chalmers (2006) to investigate the influence of attention-modulating tasks during movement on the artifacts induced by VRS foveated graphics. Our focus is on understanding the extent to which different types of movement (active versus implied) and types of tasks (a simple task involving only fixation of targets versus a more demanding task involving fixation, discrimination between, and counting of appropriate targets) impact an individual's ability to detect degradations in quality.

3 MATERIALS AND METHODS

In this work, we explore the effects of task type and movement in two separate parts. The tasks will be discussed in Section 3.2. Participants either moved by walking (Active Movement - AM condition) or were stationary and observed a representation of movement (i.e. flying) through the environment (Implied Movement - IM condition). We used NVIDIA VRS in its 4×4 configuration (1/16 of the sampling rate of the non-degraded area, which is the most aggressive setting in our system) so that we obtain maximum benefits from our method and the down-sampling artifact can be observed more clearly by the participants. We choose to use two foveation regions which we term HQ (high quality, rendered at the native shading rate - 1×1 pixel sampling) and LQ (low quality, rendered at VRS 4×4). The dynamic foveated rendering algorithm we use was implemented by adapting HTC Vive code for foveated rendering (ViveSoftware, 2020) so it functions with our current platform and XR SDK.

3.1 Apparatus

The data for this experiment were collected in the Virtual Reality Research (VR2) Facility at the University of Manchester. We used the Oculus Quest Pro headset with eye-tracking enabled (refresh rate: 72Hz). The headset has 1800x1920px resolution per eye and a rendered horizontal FOV of 108°. Because the participants had to walk through the environment, we used the Oculus Air Link to transfer data at 90Hz and we capped the FPS at 90 on our desktop configuration (Intel i7-7700K CPU and NVIDIA RTX 3080 Ti GPU). Note that HMDs tend to vary in effective resolution over the FOV because of the optical properties of the lens. Beams et al. (2020) showed that effective resolution in VR HMDs decreases rapidly for off-axes angles. This could potentially induce a confound in our findings, however studies that have reported this were done using headsets with fresnel-type lenses. The Quest Pro used in our study has pancake lenses which improve optical quality edge-to-edge (Xiong et al., 2021). Moreover, the purpose of this experiment is not to report a specific compression or foveated rendering algorithm, but to study and disentangle the interaction between the experimental conditions. Given the technology used was the same across participants, any device-specific differences should not influence the effects studied here.

The experiment was coded in the *Unity* game engine (Version 2022.3.4f1) and verbose data about each trial was collected using the Unity Experiment Framework (UXF), which was designed for the development and control of psychophysical studies (Brookes et al., 2020).

3.2 Tasks

To explore how increased attention, caused by task difficulty, can be studied using dynamic foveated rendering and in conjunction with different movement conditions, we devise a novel method in VR. The rapid serial visual presentation (RSVP) task from Huang and Dobkins (2005) consists of a series of rapidly presented random letters (i.e. distractor letters) at a fixation point, with participants tasked with finding a target letter. Increasing the number of distractor letters also increases the task difficulty and therefore the level of attention participants are using to complete the task. Krajancich et al. (2023) used this method successfully to explore the effects of attention on foveated graphics.

Inspired by this, we propose two tasks: *Simple Fixations (SF)* and *fixation, discrimination, and counting*, which we will refer to as the *Monkey Finder (MF)* task. We divided the Sponza Scene corridor into four distinct spawn areas. Note that the y-axis position of the spawn volume is scaled to the height of the participant at the beginning of the experiment. In both tasks, participants move through the scene and are instructed to look at appearing objects. Objects appear sequentially before the participant enters each spawn zone so that items always appear within the FOV of the viewing frustum. When an object is intercepted by the gaze of the participant, it blinks once and disappears. Objects spawn randomly within the volume of the area (1.25×1×1 meters). Each corresponding trial in each task condition was allocated the same positional values at runtime (e.g. items in trial 1 in MF had the same positions as in trial 1 in SF). This gave us confidence that the participants performed roughly the same eye fixations in both tasks. At the end of each SF and MF trial participants were asked if they noticed a degradation. To provide a reference, participants always saw the undegraded scene (i.e. native shading, VRS1×1) prior to movement beginning.

3.2.1 Simple Fixations (SF)

In the SF condition, participants move through the environment and are presented with spheres (top-middle in Figure 2). Participants are instructed to fixate on spheres as they progress through the environment. Note, in both tasks we used textureless objects while randomizing the color of each spawned item.

3.2.2 Monkey Finder (MF)

In this task, participants move through the environment (AM and IM). They are instructed to fixate on items that appear sequentially and to count the occurrences of a target object (here, *Suzanne, the Blender Monkey*, top-left in Figure 2). Compared to SF, the objects that appear on the screen can now be any of the well-known models in the Graphics community presented in Figure 2. The probability of a monkey appearing was set to 33%, compared to 13.2% for the distractor objects. This was to ensure participants were counting and staying engaged with the task. At the end of the trial, participants were asked how many monkeys they spotted (between 0-4) in addition to the degradation question.

3.3 Stimuli

We adapted a model of the popular Crytek Sponza Scene rendered in Unity (Universal Render Pipeline) and scaled it to match the dimensions of the research facility. The facility has a length of 5.5m; our path was capped at 4.5m (Figure 1, right). We turned all the anti-aliasing and lighting probes provided by Unity off in order to not induce any confounds.

We added a semitransparent guiding sphere that occupied 2° of visual angle in front of the partici-

Figure 1: Left: representation of VRS4×4 shading reduction; in green the foveated region (HQ); the middle dot represents the eye-tracked focal point. The black margins show the algorithm culling everything outside the peripheral FOV. Towards right: a comparison to native shading (VRS1×1). Right: the path participants walked scaled to the dimension of the facility.

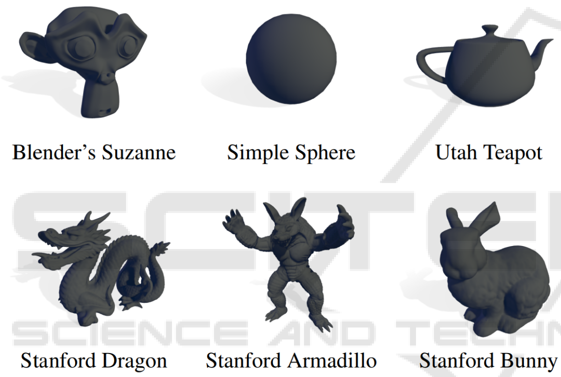

Figure 2: Popular Computer Graphics models used for MF and SF tasks.

pants and served as a guide for the speed they had to achieve. The sphere accelerated to 0.75m/s in 0.25s in order to mimic the acceleration of a human initiating walking. This speed was chosen to match typical walking speeds in VR, which are slower than real locomotion (Perrin et al., 2019). When the sphere reached its peak velocity, the VRS degradation was triggered.

In each trial, participants were informed whether they were performing SF or MF (interleaved) and then initiated movement. At the end of each trial, they were asked if they noticed any degradation. They were shown what the VRS degradation looked like in the practice trials before the experiment. At the end of the trial, participants were presented with an answer screen. If they had performed the MF task they were asked how many monkeys they counted and whether they noticed any degradation by pressing either the *yes* or *no* button. In the walking condition, they were asked to rotate and press the trigger on the controller to start the next trial. In the passive condition, they simply pressed the trigger to initiate the next trial.

3.4 Design

We used a 2×2 fully factorial within-subjects design for a total of four experimental conditions. Our independent variables were type of movement (AM versus IM) and task type (SF versus MF). We compared performance on our dependent variable (i.e. the probability of participants' not noticing the quality degradation) between conditions. For each trial, the diameter of the HQ region (Figure 1, left) was varied using a Kesten staircase. The diameter varied in line with two interleaved, partially overlapping adaptive staircases for each of the four experimental blocks. One staircase started with a large diameter (i.e. HQ = 90°), and the other started with zero diameter (i.e. full degradation, HQ = 0°). For more information about the Kesten staircase, see Section 3.5. Each staircase was comprised of 40 trials. To ensure participant engagement we also added catch trials. In a catch trial, we presented a VRS4×4 degradation where HQ diameter = 0°, and all of the spawned objects were spheres that appeared in the center of the visual field. We presented eight catch trials per experimental condition in a randomized order, resulting in a total of 336 trials per participant across the four conditions. In order to minimize fatigue, we ran the experiment in two sessions; one in which participants undertook trials involving the AM condition (with SF and MF) by producing self-induced surge motion (i.e. forward walking), and one in which they were stationary and were moved through the scene (the IM condition with SF and MF). We formulated three hypotheses for our experiment which were pre-registered with the Open Science Framework (OSF) at https://osf.io/wkptz. Note that we retrospectively changed the pre-registered term 'cognitive load' to 'attentional load' in this paper. We do not deviate from our pre-registered analysis plans.

- H1: Virtual Reality users engaged in tasks that require a higher attentional load will tolerate higher levels of degradation in a foveated system compared to those users not engaged in high cognitive load tasks.
- H2: Users engaged in Active Movement will tolerate a significantly higher level of degradation compared to users engaged in Implied Movement.
- H3: Users who are both walking and performing higher attentional load tasks will tolerate greater levels of degradation compared to users subject to all other conditions/combinations of conditions.

3.5 Kesten Staircase

The *Kesten Adaptive Staircase*, or *Accelerated Stochastic Approximation* is an algorithm used in psychophysics that has the advantage of converging rapidly towards a threshold when compared to many other adaptive staircase procedures (Treutwein, 1995). In our case, we set the staircase to converge at two symmetrical values around the 75% performance threshold for VRS degradation (i.e. the diameter that gives rise to a 75% chance of not detecting the degradation). This means they were three times more likely not to notice the change than to notice it.

$$diam(HQ) = x_{n+1} = x_n - \frac{step}{2 + m_{revs}}(resp_n - \Phi), n > 2 \quad (1)$$

Equation 1 represents the accelerated stochastic approximation. For the first two trials, this does not differ from the standard stochastic staircase procedure (i.e. the reversals in the response category are not accounted for). The value of $diam(HQ)$ represents the diameter of the HQ region that is manipulated in each trial using the result obtained from the staircase procedure. The variable m_{revs} is a cumulative count of the number of times participants reversed their responses. The variable $resp_n$ is a binary value (*yes = 0* or *no = 1* in our case) representing the answer the participant gave about whether they noticed the degradation for the n^{th} (i.e. the previous) trial in the staircase. We use two interleaved staircases (one ascending towards the threshold from HQ = 0° upwards, and one descending from HQ = 90° downwards). This generates more data around the performance point of interest (here 75%). We chose $x_0 = 0, \phi = 0.875$ - ascending; $x_0 = 0.9, \phi = 0.625$ - descending. The $step$ value was set at 0.45. This value controls the initial step size, which is reduced as a function of the number of reversals.

3.6 Participants

We collected data from 15 participants (12 naive to the study, and three involved with designing the experiment). All participants were staff members at the University of Manchester. Data from two participants were discarded because their responses did not converge and it was therefore not possible to fit psychometric functions to their data, and from one because they did not complete both parts of the experiment. The study was approved by the Ethics Committee of The University of Manchester, Division of Neuroscience and Experimental Psychology. Written consent was given by all participants. The AM session lasted ≈ 40 minutes and the IM ≈ 25 minutes. The effective time of the experiment was between 60-70 minutes per participant with breaks offered every 20 minutes in order to prevent fatigue and preserve data quality. Eye-tracking data was collected from all participants. Note, we added an additional eye-tracking calibration test at the beginning of the experiment in order to make sure there were no deviations caused by the Quest Pro calibration.

3.7 Pre-Screening

In order to confirm task comprehension and allow participants time to acclimatize to the VR paradigm, they were given pre-experimental practice trials in each condition until they felt comfortable with the task. Additionally, whilst stationary, we presented VRS degradations with HQ = 0° (full degradation) and HQ = 5° (severe degradation) and asked them to compare it to the scene rendered at 1×1 shading (native shading, no degradation).

3.8 Procedure

Each participant completed the experiment in two separate sessions. In one of the sessions they performed AM (SF and MF tasks interleaved) and in the other IM (SF and MF tasks interleaved). Participants were randomly assigned either AM or IM as their first experimental condition to prevent order and learning effects. At the beginning of each session, we calibrated the eye-tracking with the Oculus application and additional manual calibration if required. After a countdown and a text message informing the participants that they were doing MF or SF, they initiated movement. In the AM condition, they were instructed to match the speed of the fixation sphere, with the sphere changing color if they were not fast enough. At the end of an AM trial, participants answered the task-specific question and then were asked to rotate 180° until aligned with a straight line that appeared in the VR environment. At the beginning of each trial, the scene was rotated to match the midsaggital plane of the participant in order to account for small deviations in participant orientations and to present a consistent representation of the Sponza environment. In the IM trials, participants were stationary, but we aligned the scene in each trial to account for small movements.

3.9 Validation

The catch trial results show that, across conditions, participants could correctly spot VRS degradations. Eight participants correctly identified the degradation across all catch trials, and four only responded incorrectly to 1 catch trial out of 16 (total number across conditions). In order to make sure the participants completed the tasks correctly, we recorded if participants managed to look at all the objects in the scene across all trials. Across participants, the average number of objects that were gazed at is 3.79 out of 4 (total number of objects spawned per trial). This gave us confidence that participants performed the task as instructed. Moreover, we recorded the number of correct answers in the MF task. There was no significant difference between the proportion of correct answers in AM (93%) and IM (94%) MF conditions ($t(22.144) = 0.30, p = 0.75$) which provides evidence that the task was performed as instructed in both conditions.

3.10 Psychometric Functions

We fit the binary response data in our experiment using a psychometric function fitting approach. Psychometric functions (PF) model how perception relates to variations in a physical stimulus; here, the varying diameter of the HQ region. This physical quantity is then mapped to the perceptual quantity, i.e. the probability of the participant not noticing the shading reduction at each foveation level. We used a cumulative Gaussian PF with two free parameters – mean and standard deviation. Example fits can be seen in Figure 3.

The data were fit using the *quickpsy* package (Linares and López-Moliner (2016) in R (Version 4.2.3, R Core Team, 2023). This package provides functions that fit the PF to binary response data using a maximum-likelihood approach. From this fit, we recovered the diameter at which participants had a 75% probability of not noticing the degradation. Consequently, we recover the point at which participants are three times more likely to not notice the degradation than to notice the degradation. We refer to this point as *maximum tolerated diameter* (MTD). Note that in this experiment, lower MTD values for the HQ region mean more scope for optimization.

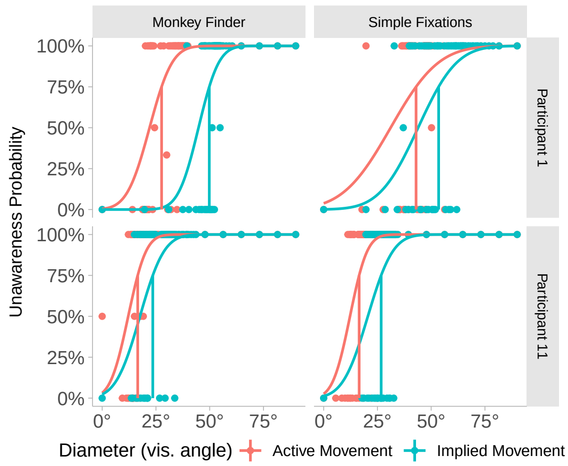

Figure 3: Example psychometric function fits for two participants.

4 RESULTS

Having satisfied the Kolmogorov-Smirnov test for normality, we use ANOVA to analyze our MTD data. We plot the mean value of the MTD points (75% likely to notice a degradation) for our dataset in Figure 4. Here, smaller MTD values indicate tolerance for a larger degraded area, increasing the scope for optimization. We observe that the mean thresholds span from 34.3° to 51.3° in all conditions (see Table 2). This shows that even in a very aggressive VRS configuration, there is significant scope to reduce the shading quality. Figure 4 suggests that both MF and AM manipulations lead to a potential for degrading the scene over SF and IM conditions respectively. In support of these observations, we find significant main effects of these factors in a 2×2 repeated measures ANOVA (Table 1). As such, the MTD is modulated by a main effect of Task ($F(1,11) = 7.000, p = 0.023$) and a main effect of the type of movement ($F(1,11) = 14.554, p = 0.003$). However, we do not find a significant interaction between the two main effects ($F(1,11) = 0.028, p = 0.870$) suggesting that their contributions are additive.

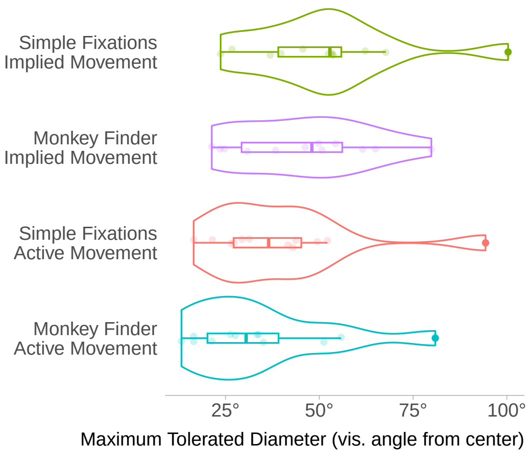

Figure 4: Violin plots showing descriptive statistics of all MTD points across all combinations of conditions (left).

Table 1: Summary statistics for the ANOVA analysis.

Conditions	F stat	*p*-value
Task	7.000	0.023*
Type of Movement	14.554	0.003*
TypeMov * Task	0.028	0.870

The results obtained here support our hypotheses. We showed that:

- A task that requires more attention causes participants to be significantly less likely to notice degradations.
- When performing active movement, participants are willing to tolerate a significantly smaller HQ region.
- The mean MDT value obtained for the MF-AM task is lower than for all other conditions, indicating that the combination of active movement and the presence of a high attentional load task is additive in nature regarding participants' tolerance to foveation.

5 DISCUSSION

The findings presented here have many potential implications for the design of future foveated graphics implementations. In consumer applications, users are often engaged in visual search and discrimination tasks (e.g. finding items, discriminating between allies and enemies). Moreover, the rise of VR arcades and the increasing popularity of Mixed Reality (MR) headsets suggests that we are moving away from the general paradigm of limited space, at-home VR, and that self-locomotion will become more common. Understanding how different types of motion

Table 2: Mean MTD values across the sample for all combinations of conditions.

	AM	IM
MF	34.3°	45.5°
SF	39.7°	51.3°

affect foveated graphics will be crucial in these future VR and MR systems. Our study is unique in that we combine types of movement and tasks, as both are important aspects of immersive media. We provide a novel method for analyzing their combination during forward surge movements. It is important to mention that in this experiment, we use a state-of-the-art shading method and a fully dynamic rendering implementation, and whilst we do report potential savings, we do not claim a fully functional foveated algorithm based on the conditions explored here.

Comparing our findings with those in Petrescu et al. (2023a), we suggest that when using our method and accounting for both task and type of movement, the HQ region only needs to occupy 31.7% of the rendered FOV in the HMD in the 4×4 configuration, compared to their 38.8%. Moreover, we show that this effect holds for a fully foveated system without participant gaze-restraints. Ellis and Chalmers (2006) created a similar model that scaled the HQ region to the hypothetical force experienced by the vestibular system. They created a dynamic map of ego-movement that guided sampling and found an HQ region of 65% of the FOV. It is important to mention, however, that they used a downsampling method of a much older renderer, rather than a VRS-like solution. We believe that our system would benefit from a dynamic scaling of the HQ area, such as that presented by them.

Task type has been shown to influence the visual behaviors of the participant. Malpica et al. (2023) showed that during visual search in VR (a similar paradigm to ours), participants exhibit significant differences in visual behavior resulting in larger saccades and shorter fixations during scene analysis. Perhaps, our results indicate that visual artifacts generated by VRS could be masked due to these movements. However, in the study mentioned above, participants are moved through the environment (implied movement) at a steady pace and do not produce self-induced locomotion. Lisboa et al. (2023) showed that increasing the velocity of the user in instances of implied movement significantly reduces their ability to detect peripheral degradations in a foveated system. We only use one velocity in our experiments, but their findings motivate further work in which multiple velocities for active movement are explored using a motion platform. Moreover, we provide evidence that the addition of difficult tasks (which require more attention) represents an untapped resource with regards to maximizing foveation benefits. This is supported by the findings in Krajancich et al. (2023), which showed that tasks which require more attentional focus result in lower sensitivity to peripheral loss of detail. During active movement, vestibular, visual, proprioceptive, and efference copy systems are presented with consistent input and interact in complex ways (Holten and MacNeilage, 2018). This interplay might affect the way users perceive foveated rendering degradations. Our claim is further supported by the fact that retinal motion is minimized during forward surge in order to stabilize information about depth (Liao et al., 2010), which could also stabilize temporal aliasing induced by shading reduction.

Limitations and Future Work: Firstly, the eye tracker that comes with the Quest Pro only updates at 72Hz. There is evidence that critical fusion frequency is between 50-90Hz, but may be as high as 500Hz (Mankowska et al., 2021). We observed that in some cases, large saccades caused a delay in the HQ region being updated based on the gaze. Secondly, participants might have noticed the popping effect of the degradation after achieving the desired velocity. This will be addressed in future work and we believe a dynamic scaling such as the one employed by Ellis and Chalmers (2006) could be used to solve this. Third, we only use two layers of foveation. Most state-of-the-art algorithms benefit from at least three foveation and/or blending layers. Adding additional layers would have created a combinatorial explosion of trials. Unlike the RSVP task found in Krajancich et al. (2023), our task does not scale in a convenient manner. In order to be able to fit a model of our effect, we need more data points across levels of attentional load. In our case, the experiment was already very long so adding an extra dimension to a within-participants study would be time consuming and expensive to run.

The results of exploring the coarse perceptual effects of attention and movement in the present study raise important questions. While we only studied simple visual search, there is reason to believe that other visual behaviors, such as those studied in Malpica et al. (2023), may yield different results regarding task performance and attentional load. The study of the effects of type of motion on peripheral acuity would also benefit from an investigation using low-level visual stimuli, such as Gabor patches; doing so would allow for the collection of baseline data that would be able to inform vision models such as those found in Mantiuk et al. (2021) or Mantiuk et al. (2022). Decoupling the effects of the vestibular and visual system (e.g. through the use of motion platforms) could also give more insight into what causes AM to mask more degradation.

6 CONCLUSION

In this paper, we examined if artifacts generated by shading reductions with VRS4$\times$4 in a dynamic foveated rendering paradigm can be masked by task and type of movement. We find that active walking and more demanding tasks increase participants' tolerance to such degradations. Given the push towards VR and AR standalone devices, we expect users to move more whilst performing visual searches of GUIs or tasks in entertainment applications. We hope these findings will inspire more research that aims to understand how these mechanisms interact and how they can be implemented in existing models or applications.

ACKNOWLEDGEMENTS

This research has been funded and supported by UKRI. We thank Boris Otkhmezuri from The University of Manchester, Virtual Reality Research (VR2) Facility, for his valuable insights.

REFERENCES

Beams, R., Collins, B., Kim, A. S., and Badano, A. (2020). Angular Dependence of the Spatial Resolution in Virtual Reality Displays. In *2020 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 836–841. ISSN: 2642-5254.

Braun, D. I., Schütz, A. C., and Gegenfurtner, K. R. (2017). Visual sensitivity for luminance and chromatic stimuli during the execution of smooth pursuit and saccadic eye movements. *Vision Research*, 136:57–69.

Brookes, J., Warburton, M., Alghadier, M., Mon-Williams, M., and Mushtaq, F. (2020). Studying human behavior with virtual reality: The Unity Experiment Framework. *Behavior Research Methods*, 52(2):455–463.

Cave, K. R. and Bichot, N. P. (1999). Visuospatial attention: Beyond a spotlight model. *Psychonomic Bulletin & Review*, 6(2):204–223.

Curcio, C. A. and Allen, K. A. (1990). Topography of ganglion cells in human retina. *Journal of Comparative Neurology*, 300(1):5–25. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1002/cne.903000103.

Denes, G., Jindal, A., Mikhailiuk, A., and Mantiuk, R. K. (2020). A perceptual model of motion quality for rendering with adaptive refresh-rate and resolution. *ACM Transactions on Graphics*, 39(4).

Driver, J. and Baylis, G. C. (1989). Movement and visual attention: the spotlight metaphor breaks down. *Journal of Experimental Psychology. Human Perception and Performance*, 15(3):448–456.

Duinkharjav, B., Chen, K., Tyagi, A., He, J., Zhu, Y., and Sun, Q. (2022). Color-Perception-Guided Display Power Reduction for Virtual Reality. *ACM Transactions on Graphics*, 41(6):210:1–210:16.

E. Jacobs, D., Gallo, O., A. Cooper, E., Pulli, K., and Levoy, M. (2015). Simulating the Visual Experience of Very Bright and Very Dark Scenes. *ACM Transactions on Graphics*, 34(3):25:1–25:15.

Ellis, G. and Chalmers, A. (2006). The effect of translational ego-motion on the perception of high fidelity animations. *Proceedings - SCCG 2006: 22nd Spring Conference on Computer Graphics*.

Guenter, B., Finch, M., Drucker, S., Tan, D., and Snyder, J. (2012). Foveated 3D graphics. *ACM Transactions on Graphics*, 31(6):164:1–164:10.

Harley, D. (2020). Palmer Luckey and the rise of contemporary virtual reality. *Convergence*, 26(5-6):1144–1158. Publisher: SAGE Publications Ltd.

Holten, V. and MacNeilage, P. R. (2018). Optic flow detection is not influenced by visual-vestibular congruency. *PLoS ONE*, 13(1):e0191693.

Huang, L. and Dobkins, K. R. (2005). Attentional effects on contrast discrimination in humans: evidence for both contrast gain and response gain. *Vision Research*, 45(9):1201–1212.

Jindal, A., Wolski, K., Myszkowski, K., and Mantiuk, R. K. (2021). Perceptual model for adaptive local shading and refresh rate. *ACM Transactions on Graphics*, 40(6):281:1–281:18.

Kandel, E., Schwartz, J., Jessell, T., Jessell, D. o. B. a. M. B. T., Siegelbaum, S., and Hudspeth, A. J. (2012). *Principles of Neural Science, Fifth Edition.* McGraw-Hill Publishing, Blacklick. OCLC: 1027191624.

Krajancich, B., Kellnhofer, P., and Wetzstein, G. (2021). A perceptual model for eccentricity-dependent spatio-temporal flicker fusion and its applications to foveated graphics. *ACM Transactions on Graphics*, 40(4):47:1–47:11.

Krajancich, B., Kellnhofer, P., and Wetzstein, G. (2023). Towards Attention–aware Foveated Rendering. *ACM Transactions on Graphics*, 42(4):77:1–77:10.

Liao, K., Walker, M. F., Joshi, A. C., Reschke, M., Strupp, M., Wagner, J., and Leigh, R. J. (2010). The linear vestibulo-ocular reflex, locomotion and falls in neurological disorders. *Restorative Neurology and Neuroscience*, 28(1):91–103. Publisher: IOS Press.

Linares, D. and López-Moliner, J. (2016). quickpsy: An R Package to Fit Psychometric Functions for Multiple Groups. *The R Journal*, 8(1):122.

Lisboa, T., Macêdo, H., Porcino, T., Oliveira, E., Trevisan, D., and Clua, E. (2023). Is Foveated Rendering Perception Affected by Users' Motion? In *2023 IEEE International Symposium on Mixed and Augmented Reality (ISMAR)*, pages 1104–1112. ISSN: 2473-0726.

Luebke, D. and Hallen, B. (2001). *Perceptually Driven Simplification for Interactive Rendering*. The Eurographics Association. Accepted: 2014-01-27T13:49:14Z ISSN: 1727-3463.

Mahjoob, M., Heravian Shandiz, J., and Anderson, A. J. (2022). The effect of mental load on psychophysical and visual evoked potential visual acuity. *Ophthalmic and Physiological Optics*, 42(3):586–593. _eprint: https://onlinelibrary.wiley.com/doi/pdf/10.1111/opo.12955.

Malpica, S., Martin, D., Serrano, A., Gutierrez, D., and Masia, B. (2023). Task-Dependent Visual Behavior in Immersive Environments: A Comparative Study of Free Exploration, Memory and Visual Search. *IEEE transactions on visualization and computer graphics*, 29(11):4417–4425.

Mankowska, N. D., Marcinkowska, A. B., Waskow, M., Sharma, R. I., Kot, J., and Winklewski, P. J. (2021). Critical Flicker Fusion Frequency: A Narrative Review. *Medicina*, 57(10):1096.

Mantiuk, R. K., Ashraf, M., and Chapiro, A. (2022). stelaCSF: a unified model of contrast sensitivity as the function of spatio-temporal frequency, eccentricity, luminance and area. *ACM Transactions on Graphics*, 41(4):1–16.

Mantiuk, R. K., Denes, G., Chapiro, A., Kaplanyan, A., Rufo, G., Bachy, R., Lian, T., and Patney, A. (2021). FovVideoVDP: a visible difference predictor for wide field-of-view video. *ACM Transactions on Graphics*, 40(4):49:1–49:19.

Murphy, B. J. (1978). Pattern thresholds for moving and stationary gratings during smooth eye movement. *Vision Research*, 18(5):521–530.

NVIDIA (2018). VRWorks - Variable Rate Shading (VRS).

Patney, A., Salvi, M., Kim, J., Kaplanyan, A., Wyman, C., Benty, N., Luebke, D., and Lefohn, A. (2016). Towards foveated rendering for gaze-tracked virtual reality. *ACM Transactions on Graphics*, 35(6):179:1–179:12.

Perrin, T., Kerhervé, H. A., Faure, C., Sorel, A., Bideau, B., and Kulpa, R. (2019). Enactive Approach to Assess Perceived Speed Error during Walking and Running in Virtual Reality. In *2019 IEEE Conference on Virtual Reality and 3D User Interfaces (VR)*, pages 622–629. ISSN: 2642-5254.

Petrescu, D., Warren, P. A., Montazeri, Z., Otkhmezuri, B., and Pettifer, S. (2023a). Foveated Walking: Translational Ego-Movement and Foveated Rendering. In *ACM Symposium on Applied Perception 2023*, SAP '23, pages 1–8, New York, NY, USA. Association for Computing Machinery.

Petrescu, D., Warren, P. A., Montazeri, Z., and Pettifer, S. (2023b). *Velocity-Based LOD Reduction in Virtual Reality: A Psychophysical Approach*. The Eurographics Association. Accepted: 2023-05-03T06:02:55Z ISSN: 1017-4656.

R Core Team (2023). *R: A Language and Environment for Statistical Computing*. R Foundation for Statistical Computing, Vienna, Austria.

Suchow, J. W. and Alvarez, G. A. (2011a). Background motion silences awareness of foreground change. In *ACM SIGGRAPH 2011 Posters*, SIGGRAPH '11, page 1, New York, NY, USA. Association for Computing Machinery.

Suchow, J. W. and Alvarez, G. A. (2011b). Motion Silences Awareness of Visual Change. *Current Biology*, 21(2):140–143.

Treutwein, B. (1995). Adaptive psychophysical procedures. *Vision Research*, 35(17):2503–2522.

Tursun, C. and Didyk, P. (2022). Perceptual Visibility Model for Temporal Contrast Changes in Periphery. *ACM Transactions on Graphics*, 42(2):20:1–20:16.

Tursun, O. T., Arabadzhiyska-Koleva, E., Wernikowski, M., Mantiuk, R., Seidel, H.-P., Myszkowski, K., and Didyk, P. (2019). Luminance-contrast-aware foveated rendering. *ACM Transactions on Graphics*, 38(4):98:1–98:14.

Vaidyanathan, K., Salvi, M., Toth, R., Foley, T., Akenine-Möller, T., Nilsson, J., Munkberg, J., Hasselgren, J., Sugihara, M., Clarberg, P., Janczak, T., and Lefohn, A. (2014). Coarse pixel shading. *High-Performance Graphics 2014, HPG 2014 - Proceedings*, pages 9–18.

ViveSoftware, H. (2020). Vive Foveated Rendering - Developer Resources.

Wang, L., Shi, X., and Liu, Y. (2023). Foveated rendering: A state-of-the-art survey. *Computational Visual Media*, 9(2):195–228.

Warren, P. A., Bell, G., and Li, Y. (2022). Investigating distortions in perceptual stability during different self-movements using virtual reality. *Perception*, page 03010066221116480. Publisher: SAGE Publications Ltd STM.

Weier, M., Roth, T., Kruijff, E., Hinkenjann, A., Pérard-Gayot, A., Slusallek, P., and Li, Y. (2016). Foveated Real-Time Ray Tracing for Head-Mounted Displays. *Computer Graphics Forum*, 35:289–298.

Xiong, J., Hsiang, E.-L., He, Z., Zhan, T., and Wu, S.-T. (2021). Augmented reality and virtual reality displays: emerging technologies and future perspectives. *Light: Science & Applications*, 10(1):216. Number: 1 Publisher: Nature Publishing Group.

Yang, L., Zhdan, D., Kilgariff, E., Lum, E. B., Zhang, Y., Johnson, M., and Rydgård, H. (2019). Visually Lossless Content and Motion Adaptive Shading in Games. *Proceedings of the ACM on Computer Graphics and Interactive Techniques*, 2(1):6:1–6:19.

Ye, J., Meng, X., Guo, D., Shang, C., Mao, H., and Yang, X. (2023). Neural Foveated Super-Resolution for Real-time VR Rendering. preprint, Preprints.