

Limits of peripheral acuity and implications for VR system design

David Hoffman (SID Member) 
Zoe Meraz
Eric Turner

Abstract — At different retinal locations, we measured the visual system’s sensitivity to a number of artifacts that can be introduced in near-eye display systems. One study examined the threshold level of downsampling that an image can sustain at different positions in the retina and found that temporally stable approaches, both blurred and aliased, were much less noticeable than temporally volatile approaches. Also, boundaries between zones of different resolution had low visibility in the periphery. We also examined the minimum duration needed for the visual system to detect a low-resolution region in an actively tracked system and found that low-resolution images presented for less than 40 ms before being replaced with a high-resolution image are unlikely to be visibly degraded. We also found that the visual system shows a rapid falloff in its ability to detect chromatic aberration in the periphery. These findings can inform the design on high performance and computationally efficient near-eye display systems.

Keywords — *foveated rendering, peripheral acuity, artifact visibility, downsampling, scintillation, flicker, latency.*

DOI # 10.1002/jsid.730

1 Introduction

The human visual system manages to perceive a wide field of view while simultaneously offering the percept of good detail at all locations. This percept is an illusion; the photoreceptor sampling is non-uniform throughout the retina, and the bandwidth of nerve fibers connecting the retina to the brain is only capable of transmitting a small fraction of the information available.¹ Using the digital imaging techniques that correspond to the sensitivities of the human visual system offers tantalizing possibilities at producing compelling imagery with a limited computational budget.

Head-mounted displays take advantage of these techniques by using foveated rendering: rendering each frame locally at a quality level that is commensurate with the visual system’s sensitivity.² Ideally, there is a potential for significant saving in processing, transmission, and optical design without any visible indication the image is not of uniform high quality. This work explores the sensitivity of the visual system to different types of image artifacts that are highly relevant to foveated rendering and near-eye display systems. These visual system limits suggest performance criteria for current and in development near-eye display systems. Some of the problems for which this work is relevant are as follows: eye tracking, display resolution, anti-aliasing, rendering, and optical characteristics.

To address these challenges with hardware, we conducted a series of experiments (summarized in Table 1) to probe the sensitivity to various types of image degradation for different

parts of the retina. In one set of experiments, we explore different rendering techniques that downsample an image. The approaches include a processing analogous to high-performance anti-aliased content (blur); a temporally stable approach that attempts to align aliases with content (stable alias); and a frame-by-frame downsampling approach that can introduce spatial as well as temporal artifacts (volatile alias). We explored the sensitivity of the visual system to these artifacts near the fovea as well as in different parts of the periphery. As part of this work, we consider the visibility of these artifacts at transition boundaries between different levels of downsampling. In another experiment, we explore the visibility of short duration exposure to a downsampled image before it is corrected to the full-resolution image, and this work can be applicable to latency requirements for an active system. We also study how the sensitivity to chromatic aberration differs as a function of eccentricity.

2 Motivation and background

Virtual reality (VR) as a platform has several fundamental differences from conventional displays. The current generation of VR headsets still yield a modest field of view of 90°–100° horizontally, which although is far less than the human binocular field of view of 220°, is still far greater than the angular coverage of the normal use-case of any direct-view display. Even though VR displays feature impressive pixel counts, the perceived pixel density is still much worse than conventional monitors.^{3,4}

Received 03/28/18; accepted 07/12/18.

The authors are with the Google Inc., Mountain View, CA, USA; e-mail: hoffmandavid@google.com.

© 2018 Society for Information Display 1071-0922/18/0730\$1.00.

TABLE 1 — Summary of experiments.

Experiment	Investigation
1A	Compare the relative sensitivities to different spatiotemporal downsampling strategies.
1B	Investigate if sensitivity to downsampling artifacts shows asymmetry in the retina.
1C	Investigate if transition boundary in periphery increases visibility of downsampling.
2	Maximum presentation time of degraded imagery that is sub-threshold.
3	Quantifying sensitivity to transverse chromatic aberration in the periphery.

As a result, real-time rendering for VR has a disadvantage. It is more expensive to render a scene of equivalent sophistication as a modern video game, judged by triangle count, texture detail, and so on. VR systems need to render the scene at least twice per frame – once per eye – at a higher resolution. With limited compute power, this results in perceptibly degraded quality. Many techniques have been developed to improve the rendering efficiency in VR, such as multiview stereo rendering,^{3,6} improvements in multisampling,^{7–9} and foveated rendering.^{10–14}

Each of these methods reduces the rendering load per frame based on some visual trade-off. Foveated rendering in particular allows computational savings by reducing the total number of pixels needed to be processed. By producing lower resolution rendering in the periphery, the goal is to spend less time in areas for which the user will not notice a difference. This technique relies on the acuity limit of the human visual system, which has been measured to have a dramatic falloff with eccentricity angle.^{1,15–17} However, many of the past studies relied on static letter acuity stimuli and did not measure sensitivity to dynamic content. This paper explores motion stimuli, which can expose dynamic artifacts that are not observed with static imagery, and as will be described later, the visual system is particularly sensitive to these artifacts. The most prevalent motion artifact is a form of spatiotemporal flicker that manifests as a scintillation and is a critical challenge common to all of downsampling routines due to the visual system’s sensitivity to these high-frequency frame-by-frame subtle changes.¹⁸ Because foveated rendering increases the angular spacing of pixels in the periphery, any rasterization or upsampling artifacts also get magnified, which exacerbates the scintillation. There is also clear literature showing that motion thresholds in the periphery are lower than other spatial acuity metrics, which is best exemplified by Mckee and Nakayama.¹⁹ This work found that there was a dramatic falloff in motion acuity between 0° and 10° eccentricity and then more subtle drop between 10° and 40°.

One generic foveation technique is to render central and eccentric regions with different acuity via separate passes. So for a single frame, a “low acuity” and “high acuity” image is generated, possibly with multiple levels in between. These images are then upsampled and composited together into

the final frame.^{12,13,20} Depending on what part of the system the compositing occurs, foveated transmission can lessen bandwidth requirements for internal interfaces between the graphics processing unit and the display panel.²¹ For this approach, image artifacts get introduced at the seams between regions and during upsampling of the lower resolution regions. Artifacts at the transition boundaries between regions can be mitigated by overlapping the low- and high-resolution regions and performing alpha blending between the two images, which effectively blurs the boundary. This overlap imposes a trade-off, where a wider overlap produces a more gradual transition but also incurs greater processing overhead. A narrower overlap is less expensive but may be more detectable.

Alternative approaches exist that allow for foveated rendering to occur in a single render-pass, which also reduces the need for transition blending. Such examples include ray-tracing approaches^{10,22} or masking-based approaches.^{23,24} Masking approaches render a full-resolution image, but use depth- or stencil-culling to mask out individual pixels. Both of these approaches are less efficient than just rendering a low-resolution frame directly but can provide a visual benefit. Rasterization still requires a full-resolution image, but texture and fragment shading only occur on the non-masked pixels. Continuous transitions between levels of acuity can be formed without the need of overlapped rendering since arbitrary patterns can be generated by masking individual pixels. By using a continuous resolution reduction, artifacts caused by discontinuities in resolution are removed, resulting in fewer perceived artifacts overall.

All of the aforementioned foveated rendering methods can be used either with or without eye tracking sensors. In foveation without eye tracking, the outer edges of the VR display can have lower requirements due to lens acuity falloff, vignetting, and by the inability for users to comfortably maintain a gaze at an oblique eccentricity for prolonged periods.^{25,26} Thus, some resolution reduction that scales with eccentricity is not necessarily noticeable.^{14,27,28} The use of an eye tracker can enable a more aggressive solution in which resolution scales with retinal eccentricity, allowing for a narrower high-acuity region and more aggressive resolution reduction in the periphery. However, eye trackers can also introduce other types of motion artifacts including issues related to latency, temporally stable position errors, and temporally noisy estimates.

Regardless of method, foveation requires resampling the image and compositing images with different resolutions to fill the field of view. There are three main classes of foveated rendering degradation: blur, spatial aliasing, and spatial-temporal aliasing.

For blur, it is important to differentiate between foveated rendering and foveated transmission. For foveated transmission, we can assume that the full-resolution content has already been processed, for example, a pre-rendered

360° video or cloud rendering, and foveation occurs to aid in bandwidth limits for transmission. In such a case, downsampling can occur with proper low-pass filtering and full anti-aliasing, which reduces flicker due to animation in the scene. The resulting low-resolution content is equivalent to an optical blur of the original, with no aliasing artifacts but without high-frequency details.

During foveated rendering, however, the savings comes from never producing the full-resolution content to begin with, meaning no opportunity exists to perform full low-pass filtering. Spatial aliasing occurs when the rasterization resolution is of lower frequency than the original resolution of the virtual content. Spatially stable artifacts may be introduced by the use of methodically reducing texture detail in the 3D scene, such as when an application forces textures to the next mipmap level.²⁹ Applications can also dynamically adjust the geometry complexity to reduce computation when objects are far away or in the peripheral vision.³⁰ Although these aliasing artifacts are present, they are aligned to the world geometry rather than the output display coordinates, meaning they do not produce flickering or scintillation effects during head movement.³¹

The third class of artifacts are temporally volatile aliasing effects. Much like spatial aliasing mentioned previously, this class of artifacts occur due to rasterization being performed at a lower resolution than the native fidelity of the virtual content. Aliasing artifacts are aligned with the pixel grid of the output display, rather than the geometry itself. This effect can be a result of having a relatively low-resolution display or by forcing the rendering to be performed at an artificially low resolution in certain parts of the screen, such as for foveated rendering. Because these aliasing artifacts are aligned to the screen coordinate system, as the user moves their head, the aliasing shifts relative to the virtual content. This relative movement causes flickering or scintillation effects, resulting in temporally volatile artifacts.

Temporal artifacts can change the rendered image from frame to frame even if there is minimal head movement. As the pose of the VR headset changes relative to the virtual content coordinate system, the pixel alignment moves with respect to the virtual content.

The perceptibility of these flicker artifacts for upsampled content also depends on the interpolation method used. Nearest-neighbor interpolation, for example, offers better local contrast retention than bilinear interpolation or bicubic but also amplifies these motion flicker artifacts. Similarly, kernel-based methods can minimize the visibility of static aliasing artifacts, but likely are not hardware supported, and yield further reduction in local contrast. Foveation techniques can also reduce the initial generation of scintillation artifacts prior to rendering by either randomizing the rasterization sampling³² or by aligning the pixel grid to world coordinates.³¹ Each of these methods introduce trade-offs of extra computation for improved visual quality. In the experiments described

here, the goal is to look at the broad classes of solutions rather than the efficacy of the variants within each broader category.

3 General methods

To evaluate the limits of visual sensitivity to different artifacts as a function of retinal eccentricity, we use a desktop-mounted display with a head restraint. Such a system has the benefit of a high-resolution display without degradation from an optical element. For these tests, we used an RGB stripe Sony BVM 300 4K OLED monitor. The viewing distance was held at 55 cm with a chin and forehead rest so that the full screen subtended about 60° horizontally and pixels normal to the eye subtended about 1 arcmin. The color was adjusted to Rec. 709, with 2.2 gamma and peak luminance of 130nits.

One consideration in the presentation of imagery was that the angular subtense of pixels would vary depending on the eccentricity being tested. Pixels N° off-axis for the display would have an effective angular size of $\cos^2(N)$ in the horizontal direction and $\cos(N)$ in the vertical direction. Although this could be compensated in software, this would require an additional resampling stage that could introduce additional artifacts. At 40° viewing angle, the pixels could be as little as 58% of the nominal pixel size, and this was deemed an unacceptable error. To mitigate this issue, we presented the fixation target to the left of the front surface normal of the display equal and opposite to the largest eccentricity tested (Fig. 1). Thus, at the maximum eccentricity tested, 40°, the 0° and 40° conditions were presented 20° from the front surface normal of the display and limited pixel foreshortening effects to at most 12%, which was deemed acceptable.

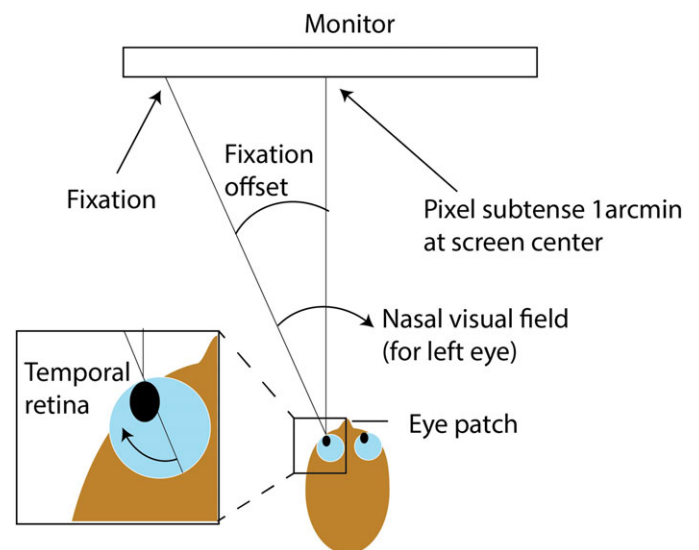


FIGURE 1 — Top view of experimental setup. Fixation is offset from the monitor center to mitigate pixel foreshortening.

3.1 Participants

Five observers, with a letter acuity of 20/20 or corrected to 20/20, completed each test. Ages ranged from 25 to 40 years old. Two of the observers were completely naive to the hypotheses and the rendering methods.

4 Experiment 1

A core question to understand the differential acuity of the visual system is determining what types of details the visual system is sensitive to at any given location of the retina. To do this, we needed to probe visual acuity with high spatial specificity and with relevant imagery for a VR system. Furthermore, motion is inherent in VR systems because even in a static scene, to avoid simulator sickness in VR imagery, objects should be counter-shifted with head rotation to resolve visual-vestibular conflict. In this experiment, we use a head restraint and a desktop monitor but include motion on the screen to ensure that we can properly showcase temporal issues in foveated rendering. We apply a circular orbit to an otherwise still image in which the image translates in a circular path with 30-arcmin diameter and 1-s period without change to its orientation.

Experiment 1 is divided up into three parts. The first part considers different types of foveated rendering algorithms. The second part considers the symmetry of visual sensitivity across the retina. The third part considers the visibility of the boundary between regions of different resolution.

4.1 Experiment 1A

In Experiment 1A, we probe three types of image artifacts at five different locations in the retina and two different images.

The types of image impairments considered were as follows:

1. blurring the image to remove the high spatial frequencies;
2. downsampling the image before shifting so that the re-sampling artifacts are static (stable alias);
3. downsampling after motion leading to resampling on a per frame basis (volatile alias).

The images to which these impairments are applied are shown in the top panels of Fig. 2. They represent a detail-heavy photograph (Crowd) and a low-polygon rendered scene (Forest). These two classes of imagery, photography, and computer graphics represent two of the most common types of imagery viewed in VR devices. With mobile computing, current processors make use of lower polygon counts than are found in desktop class rendering. The Forest image is one such scene that is the landing or home

screen for Google's VR environment. This image had strong edges. Some of the most common VR applications are YouTube and Google Street View in which people observe wide field-of-view imagery of real places that is captured with camera systems. To represent this type of image, we chose a photograph of a crowd. This image has high contrast with high-frequency details. These images each represent a different type of challenge for downsampling.

The second row panels show the image with a cylindrical blur applied. This is analogous to a high-end anti-aliasing routine that will remove the high-frequency details and not be subject to inserting spurious features from undersampling. The image exhibits excellent temporal stability.

The lower images use a nearest neighbor downsampling of the original image, followed by upsampling with bilinear interpolation. This type of scheme can produce a set of aliased qualities to the images including jagged edges and breaks in narrow features. In the second of the three impairments described previously, this downsampling is applied once to the reference image before being translated in its orbit motion (stable alias). This causes the breaks and blocks in the image to translate coherently with no temporal resampling artifacts. The aliases are registered to the image content and do not vary with time.

The third type of impairment uses the same downsampling routine as the second, but it is applied after the image position has been updated in its orbit. This leads to a downsampling based on a different subset of original pixels. The jaggies, blocks, and breaks will change locations frame to frame. This is illustrated at four time intervals during the orbit in Fig. 3. In the experiment, a full orbit occurs once per second, and the display updates 60 frames per second leading to a smooth orbit but with temporal scintillation at edges and features.

4.1.1 Methods

The task was a forced choice experiment with a reference and test image being offset vertically by 5° . Each image was 256 pixels (subtended 4.2° in the center) such that there was a 0.8° gap between them. The test/reference pair was shown with different horizontal shifts from a fixation target. The shifts were 0° , 10° , 20° , 30° , and 40° towards the right half of the left eye's field of view (i.e., temporal retina), and the right eye was patched. Whether the reference or test appeared on the top or bottom was randomized each trial. The observers were instructed to fixate on the center of the blue target and make a judgment of which of the images was unimpaired without making a saccade. An example test screen is illustrated in Fig. 4. When they could not tell, they were asked to guess. Based on their response, the level of downsampling/blur was staircased via two down one up procedure (with a clamped range of 2 to 30 arcmin). There were a total of 30 combinations of stimuli with five eccentricities, three types of impairments, and two different images. All the conditions were randomly interleaved, and

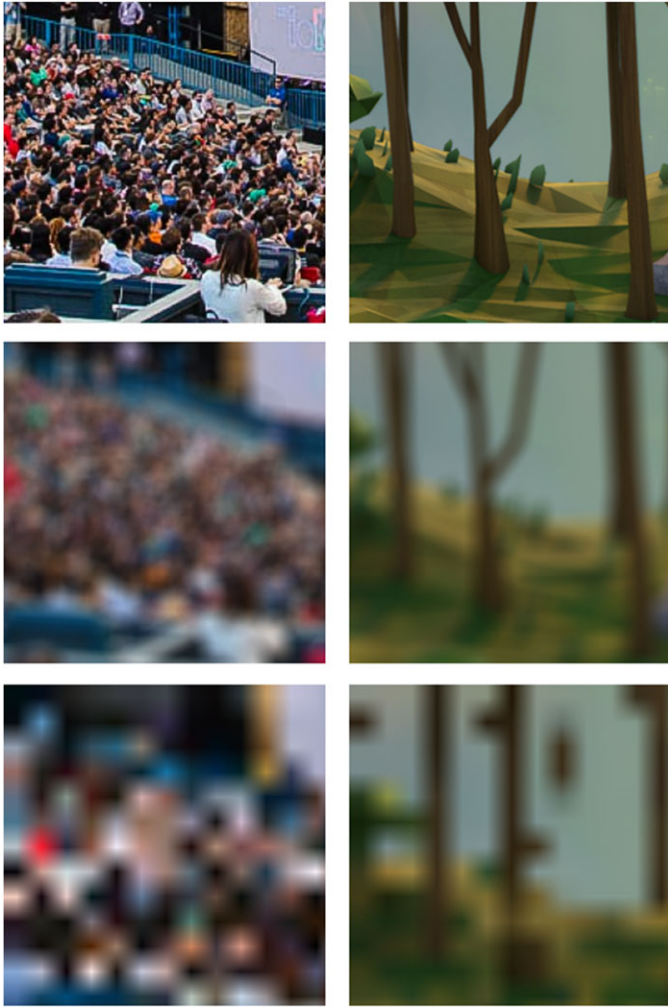


FIGURE 2 — Example illustrations of the reference images (top), blur (middle), and downsampling with bilinear upconversion (bottom). The images shown are Crowd (left) and Forest (right).

each observer completed 30 trials per condition, for a total of 900 responses.

Actual fixation was not monitored via eye tracking, but if the observer made an accidental eye movement, they were allowed to re-randomize the trial with a button press.

4.1.2 1A results

The data from each observer were fitted with a cumulative Gaussian distribution. The results are summarized in Fig. 5 as the average across observers and the standard deviation of their judgements.

There is a clear difference between the temporally volatile condition (frame downsampling) in which the aliases tended to scintillate frame by frame and the temporally stable conditions. Observers were highly sensitive to the scintillation even at large eccentricities. This is a trend that is consistent with the literature about sensitivity to motion in the periphery.¹⁹ At 40° eccentricity, observers were sensitive

to 10 arcmin downsampling in the Forest image, and 5 arcmin in the Crowd image.

There was no meaningful difference between the blur and stable alias conditions. This suggests that the observers do not discriminate between filtered and aliased high-frequency information if the information is static.

At 30° and 40° eccentricity, several observers were unable to correctly discriminate the impaired and reference images for the temporally stable conditions at the maximum impairment level tested of 30 arcmin. In these cases, we are not able to accurately estimate a threshold; in such cases, for purposes of computing an average across observers, we assume a threshold of 35, and this leads to larger error bars for the temporally stable conditions at large eccentricity.

4.2 Experiment 1B

Experiment 1A considered three impairments presented at different eccentricities that were all in the temporal retina (nasal field of view). Another important question is whether we would expect to find similar results in the nasal retina (temporal field). There has been some evidence reported that there is a nasal/temporal asymmetry that can influence letter acuity and reaction time, but it is unclear if these differences could manifest in a graphics context.^{33,34}

In this experiment, we test for whether there is a difference in peripheral acuity between $\pm 10^\circ$ and $\pm 25^\circ$ eccentricity. The experiment followed the same design as Experiment 1A, but the eccentricities tested differed, and only the temporally stabilized downsampling (stable alias) routine was tested, and the fixation offset was set to 0 such that the fixation target was at the center of the screen. Observers completed 50 trials for each condition.

4.2.1 1B results

As in Experiment 1A, we fit the response data with a cumulative Gaussian function. The average across observers is plotted in Fig. 6, with error bars representing the standard deviation. At 25° eccentricity, there was a tendency to have higher thresholds (lower sensitivity) for the temporal rather than nasal half of the retina. This trend is consistent with the contrast sensitivity experiments in the periphery.^{15,33} Other evidence in the literature such as Aubert and Forrester's work as summarized by Strasburger found that there is generally good symmetry at eccentricity less than the blind spots (at around 15° nasal retina) but some asymmetry beyond. They also note that the degree of asymmetry in the far periphery was heavily observer dependent. Our data are in agreement that at 10°, there was no consistent nasal/temporal trend with small error bars, but that at the larger eccentricities, the error bars were much larger, and there was an emerging trend towards greater sensitivity in far nasal retina than the opposite side.

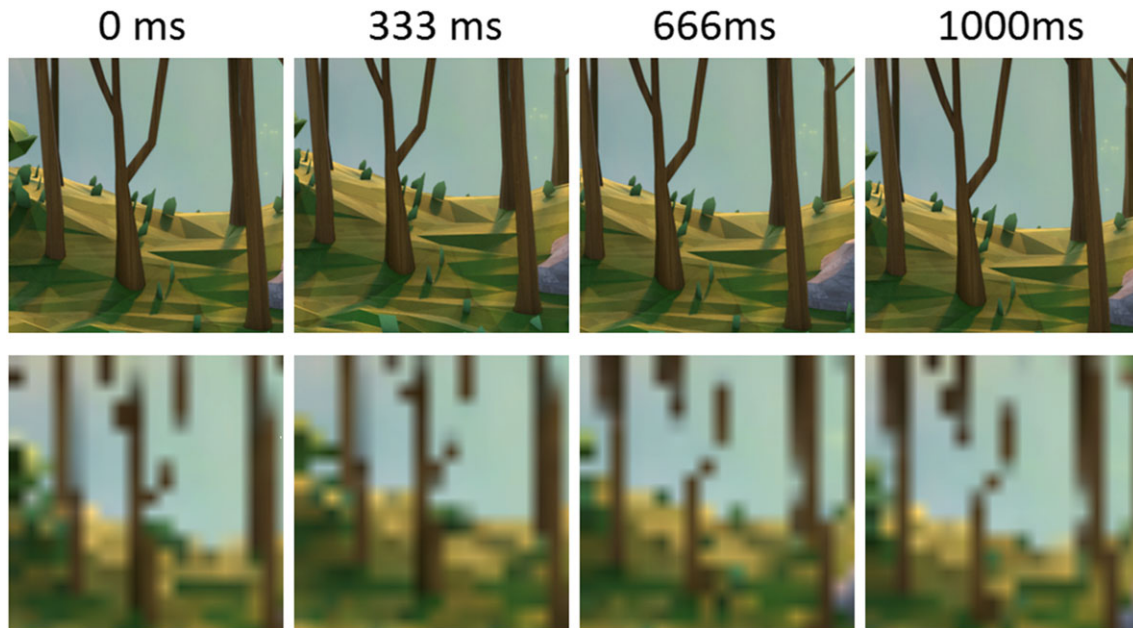


FIGURE 3 — Illustration of four phases of the orbit of the reference image and the resampled images with unstable locations of the aliasing. Top: reference image, Forest, without downsampling. Bottom: Volatile downsampling after image position update.

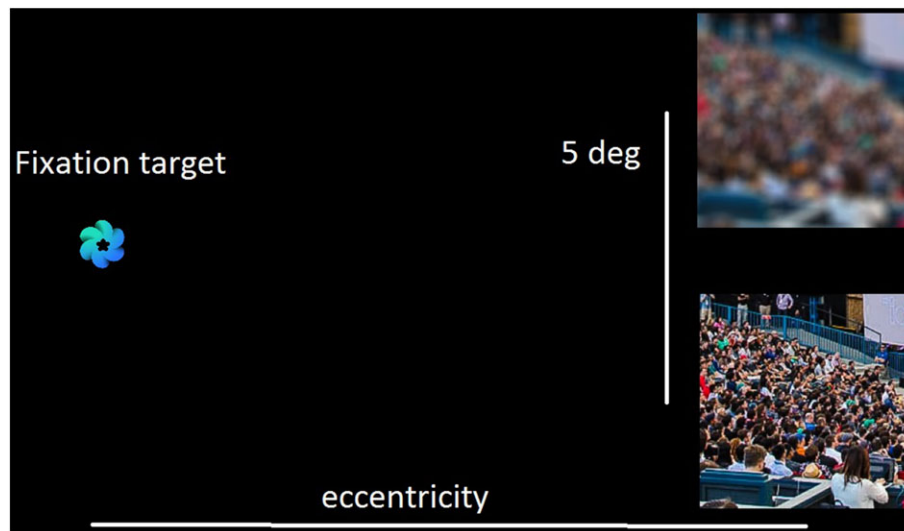


FIGURE 4 — Example presentation from Experiment 1A.

4.3 Experiment 1C

Another common issue discussed in foveated rendering is how to deal with transition boundaries between zones of different resolutions. In most implementations discussed, a blend transition is used between these regions to avoid hard discontinuities. Experiment 1C examined if a hard boundary directly exacerbates the visibility of downsampling artifacts, such as creating a temporal flicker at the boundary as objects shift from one zone to another. To evaluate the visibility of this zone, we compare the visibility of artifacts that have been uniformly downsampled and an image in

which 50% has been downsampled and the other half shown at full resolution.

The test method was the same as that in Experiment 1A, but only blur and temporally stabilized downsample conditions were tested at temporal retinal eccentricities of 10° and 20° . As before, the level of downsampling was staircased to determine observers' thresholds. These thresholds were measured for the same two images but with and without a transition dividing through the image exemplified in Fig. 7. When the transition was present, the left half of the image (closer side to fovea) was drawn without resolution reduction, whereas the right side was drawn with the reduced resolution. The location of the

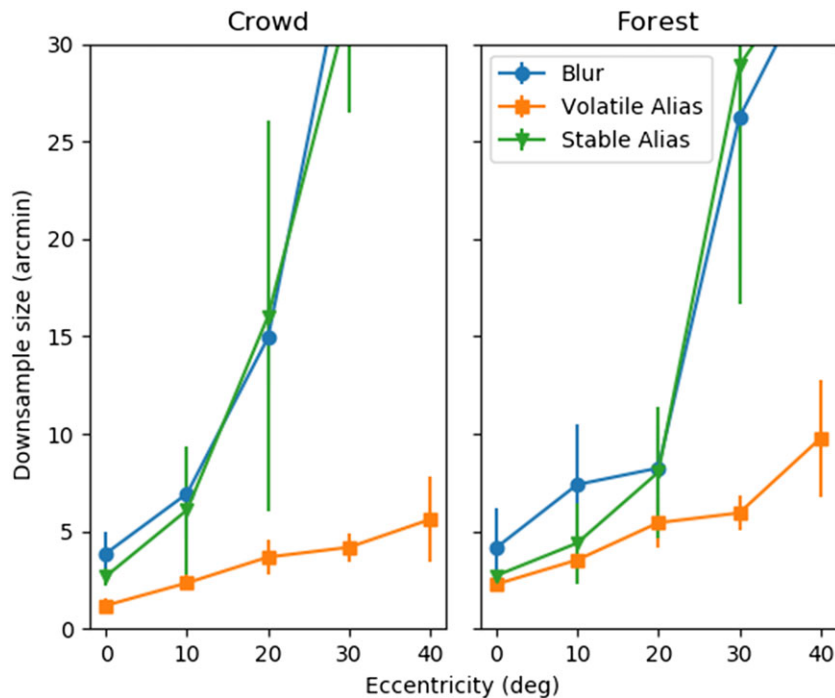


FIGURE 5 — Results of Experiment 1A. The downsampling threshold is plotted as a function of retinal eccentricity. The temporally stable strategies, blur and stable alias, are shown with blue circles and green triangles, respectively. The volatile alias condition is shown as orange squares. Tolerance for downsampling increases with eccentricity, but tolerance for temporally stable approaches increase much more dramatically.

transition remained static on the screen, while the image orbited as before.

4.3.1 1C results

If a downsampled image translating behind a boundary was enough to cause a temporal artifact, we might expect the subsampled image to have higher artifact visibility than the uniformly blurred condition. We found the opposite. The data are shown in Fig. 8. The upper figures show the results of the blur downsampling, and the lower plots show the data for the stable aliased downsampling. The left and right columns present the data from the Crowd and Forest images respectively. The blue lines represent the data from the uniform downsampling, which is analogous to Experiment 1, and the orange lines represent the split condition with partial full resolution and partial downsampled.

If a hard transition region introduced a temporal flicker, or was objectionable, we would expect that it would increase the visibility of the downsampling and thereby decrease the downsampling thresholds. We found that instead, the images with the transition regions could sustain greater downsampling before artifacts became visible.

One possible explanation is that the presence of the high-resolution imagery adjacent to the low-resolution region creates a crowding effect that makes the loss of detail in the low-resolution region less noticeable. An alternative explanation is that by making the region of the image closer

to the fovea unblurred pushes the region with blur an additional 2.4° farther into the periphery. We do not have the data points of the uniform downsampling conditions that are 2.4° farther to the fovea than the transition conditions, but based on the trends in the plots of Fig. 8, the shift hypothesis is unlikely to completely explain the difference in sensitivity. A third alternative is that by replacing half of the low-resolution image with high-resolution imagery, we have decreased the area of the low-resolution region, and this loss of area is responsible for decreased sensitivity.

5 Experiment 2: Temporal masking and latency

For dynamic foveated rendering (foveated region based on gaze estimate) to be a viable solution and be visually innocuous, the system should be able to detect an eye movement, measure the new gaze location, adjust the foveated rendering algorithm to render high-resolution imagery at the new gaze location, render the image data, transmit that data to the display, and update the pixels on the display before the visual system perceives the low-resolution imagery. Albert and colleagues conducted measurements on the minimum system latency of such a loop and found that 50–70 ms would be tolerable.³⁵

To further consider the latency requirements of foveated rendering, we developed a task to explore how the temporal

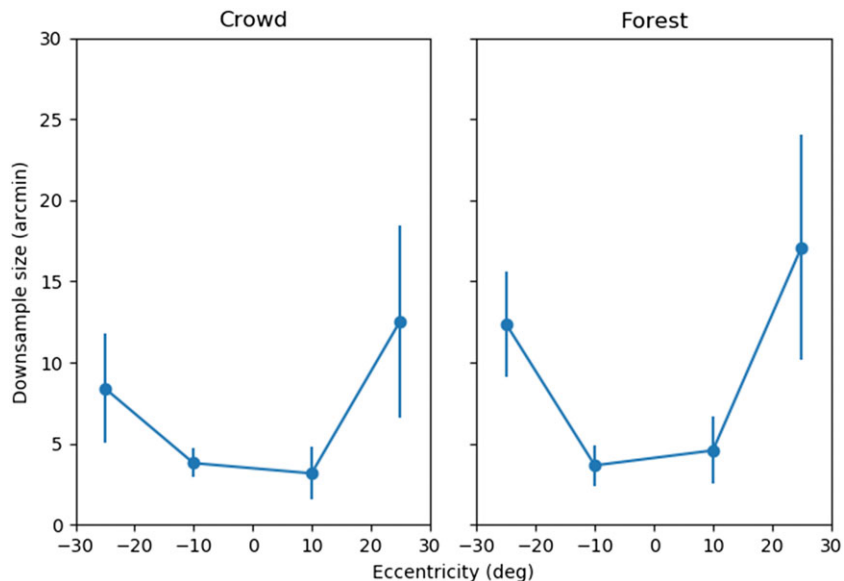


FIGURE 6 — Results from Experiment 1B. The threshold sensitivity to stable alias downsampling is plotted as a function of eccentricity. Negative eccentricities indicate nasal retina (temporal visual field), whereas positive eccentricities indicate temporal retina (nasal visual field). There is good symmetry at small eccentricities with more variability and a trend towards greater sensitivity in nasal retina at larger eccentricities.

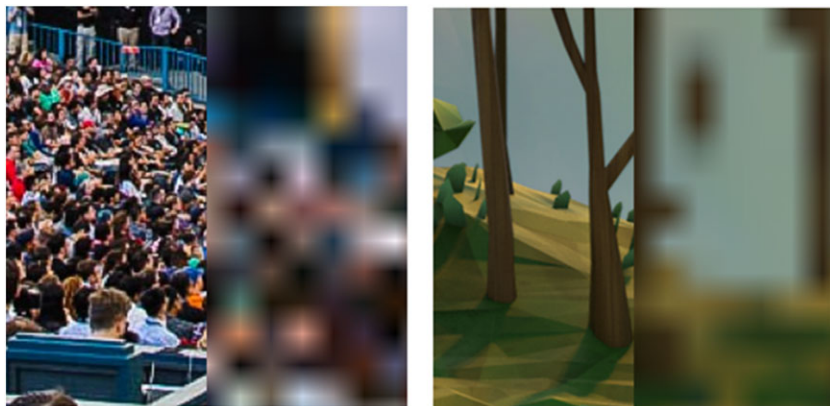


FIGURE 7 — Example stimuli showing the transition boundary between a high-resolution and low-resolution image.

masking inherent to foveated rendering could place a lower bound on the latency requirement for a dynamic foveated rendered system. In the task, there is a blank field before a sequence of images is shown in a region. The first N of these images has the downsampling operation applied before the remaining frames are shown at full resolution (illustrated in Fig. 9).

In this experimental design, saccades and system latency are irrelevant, and we are only concerned with the minimum duration a downsampled image must be shown in order for it to be perceived. The display updates at 60 Hz, and thus, each frame has a duration of 16 ms. We ask the observer after each presentation if they perceived a downsampled image before the high-resolution image. Based on their response, we adjusted the number of downsampled frames, N , until they were at 50% likelihood

of reporting that they perceived the downsampled frames. The level of downsampling was fixed at 10 and 20 arcmin using the time-stabilized alias strategy. Twelve percent of the trials were check trials in which all frames were full resolution. Each condition consisting of eccentricity, image, and downsampling level was presented 40 times. Observers were allowed to discard a trial if they were inattentive during the presentation. All judgements were made binocularly during Experiment 2.

5.1 Experiment 2 results

In Fig. 10 we plot the threshold duration ($N*16$, in ms) as a function of eccentricity. The thresholds for the 10 and 20 arcmin downsampling are shown in green and blue,

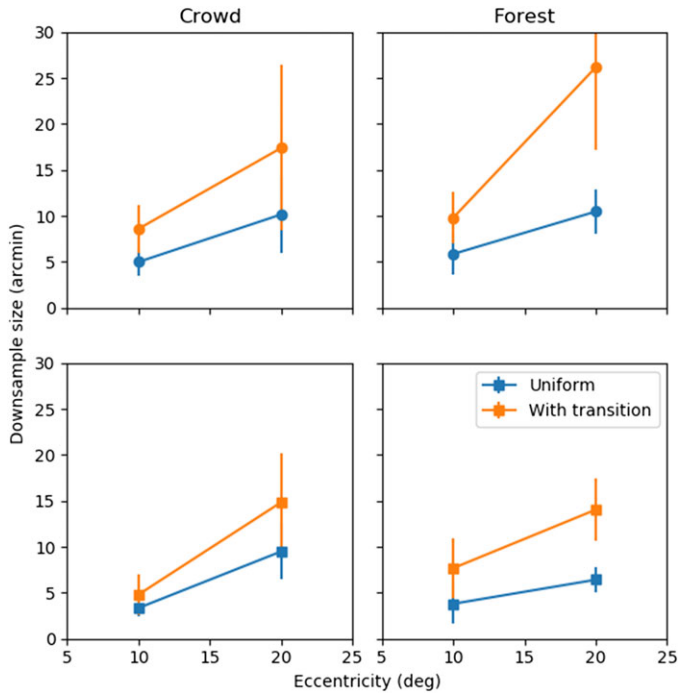


FIGURE 8 — Results from Experiment 1C. Upper plots show blur downsampling, and lower plots show stable aliasing downsampling. The image with the transition could consistently sustain more downsampling than the uniformly degraded images.

respectively. The right and left plots of Fig. 10 show the data from the Crowd and Forest images. For the five individuals tested, false positive rates on the check trials never exceeded 5%. The trends were similar for both image content with the lowest thresholds occurring at low eccentricities and larger downsampling levels.

Although we tested at larger eccentricities, the conditions and results occurring near fixation are the most informative. This offers some insight into the minimum time needed for the visual system to perceive an impaired image in the presence of subsequent masking by the full-resolution image. If the visual system makes a saccade to a location with downsampling, there will be some duration before this region can be drawn with full resolution. This measurement we make that eliminates the eye movement can offer some guidance for the latency requirements of a foveated rendering system. The findings, at both downsampling levels, that the visual system detected the downsampling if it was present for ~ 40 ms or more suggest that if the system latency for dynamic foveated rendering were at 40 ms or less, we would expect no apparent lag to be visible. This would hold for aggressive as well as subtle downsampling. In this experiment, we used a 60 Hz monitor, and it suggests that two frames of the downsampled image (32 ms) post-saccade is not visible, whereas three frames (48 ms) could be visible. In practice, with a saccade and its associated saccadic suppression, we might expect a slightly more forgiving latency requirement. These results are consistent with the conclusions of system level foveated rendering by Albert and colleagues.³⁵

6 Experiment 3

An issue of image quality that is orthogonal to downsampling but strongly related to wide field-of-view systems is chromatic aberration. Some of the headsets employ a computationally expensive operation known as chromatic aberration correction in which each of the color planes on the display are warped independently such that after passing through the lens, they appear superimposed. Other systems have optimized their optics to minimize the chromatic aberration for a central region of the view and permit some chromatic aberration in the periphery.

In this experiment, we sought to quantify the threshold magnitude of chromatic displacement at which the aberration becomes visible and characterize this for different eccentricities. We use the same method as Experiment 1, a forced choice task with two alternatives. One had registered color planes and the other had opposing horizontal offsets applied to the red and blue color planes. This offset was staircased to determine the threshold chromatic aberration at which the color displacement becomes noticeable. Example stimuli showing the simulated chromatic aberration are shown in Fig. 11.

6.1 Results

The observer responds which of the options has chromatic shift, and the separation of the blue and red color planes is adjusted by a staircase method as we used in Experiment 1. We fit the psychometric data with a cumulative Gaussian function to estimate the 75% correct point for each eccentricity. The displacement of the red and blue color planes is averaged across the five observers and shown in Fig. 12. Error bars represent the standard deviation. With eccentricity, the ability to perceive chromatic aberration drops monotonically for both images tested. At the locations near the fovea, displacements as small as 2 arcmin were perceptible. At 30° eccentricity, there was considerably more variation between observers and for the different images, but there was evidence that some displacements exceeding 20 arcmin would be perceptible.

A typical molded acrylic VR lens introduces chromatic shift of approximately 1, 8, 19, and 36 arcmin at 1° , 10° , 20° , and 30° eccentricity, respectively. In the best case scenario, this level of chromatic aberration would be at or just above the threshold if the observer was fixating at the center of the field of view (aligned with the lens optical axis). However, typical eye movements would image regions with 19 arcmin or more of chromatic shift on the fovea (0° eccentricity). Without digital chromatic aberration correction, chromatic aberrations in such a system would be clearly visible.

7 Discussion

The work we have described on measuring the limits of the visual system to resolve different artifacts across the retina

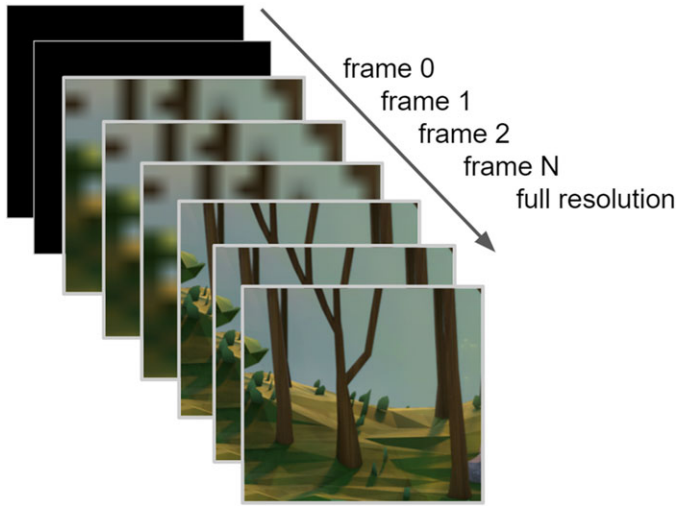


FIGURE 9 — Time course for Experiment 2, in which impaired imagery is shown for several frames preceding the full-resolution imagery.

exposes some interesting possible avenues of achieving improved image quality in VR display systems. With today's systems that have pixel sizes that are 6–8 arcmin, a number of downsampling operations could be expected to visibly

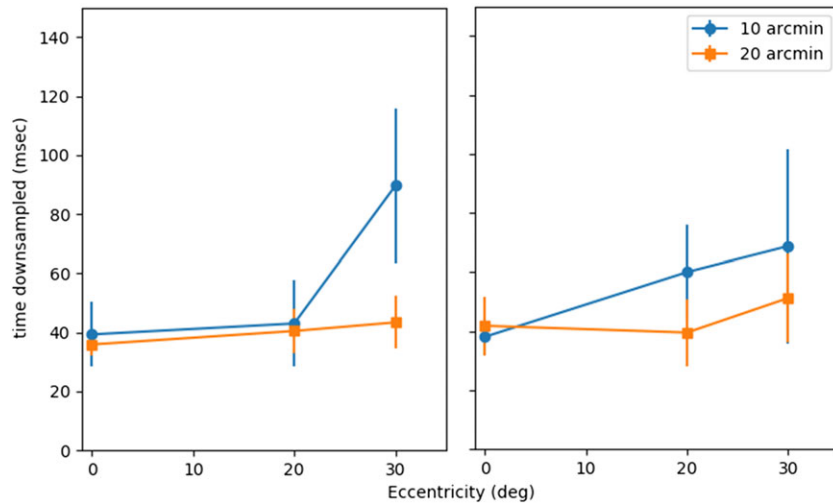


FIGURE 10 — Duration of impaired image before replacement with high quality image. The blue circles and orange squares represent 10 and 20 arcmin downsampling, respectively. Thresholds hovered near 40 ms in both downsampling conditions.

degrade the image, especially those that introduce temporal volatility. In the case of Experiment 1A, we found that with the volatile resampling, even at 40° eccentricity, downsampling to 10 arcmin spacing would be noticeable. This suggests that with current pixel sizes, it is possible that even 2 × 2 downsampling in the periphery could lead to visible scintillation.

Furthermore, at up to 30° eccentricity, the sensitivity to volatile resampling at sampling scales on the order of a single pixel (6 arcmin) indicates that anti-aliasing strategies would be likely to improve image quality by mitigating some

of the temporal fluctuation. Conversely, foveated rendering approaches that are optimized to stabilize the pixel intensities, such as world-centric downsampling, could allow sampling spacing of 15 arcmin or more at 30° eccentricity. Even with these temporally stable strategies, minimal downsampling is possible within the central 20°, while display pixel size remains at 6 arcmin or greater. As the pixel sizes shrink in next generation displays,²¹ we could expect to begin to downsample more aggressively and begin at smaller eccentricities.

Another issue we considered is the importance of the transition region between zones of different resolution. In most graphics situations, a hard boundary is highly visible and would be objectionable. For foveated rendering, this has led to most implementations using a blending region between the different zones, often with the width proportional to the eccentricity.³⁵ In our experiments, we found that a hard transition between zones in the periphery was less visible than a uniformly downsampled region; the test patches with the transition required greater blur before it reached the threshold for visibility. Earlier, we hypothesized that this effect could be related to crowding. In most other image context, the presence of extraneous details near a target decreases the

visibility of the target, and in this case, the presence of high-resolution imagery adjacent to low-resolution imagery could make the low-resolution imagery less obvious.

This finding seems to contradict the practical experience with foveated rendering that indicates that a gradual transition is needed. However, one major factor that is present in real foveated rendering systems and is absent in this experiment is eye tracking. The transition would be obvious near fixation, and thus, eye tracking failures or latency could benefit from a more gradual transition. It would ensure that at any given eye position, things are locally consistent. But, if this were the

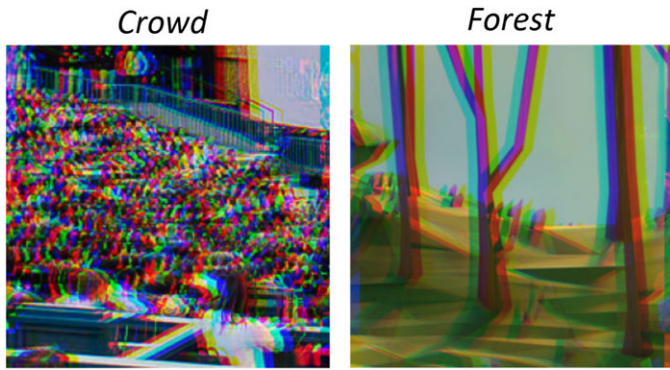


FIGURE 11 — Example stimuli from Experiment 3.

explanation, the requirement for a gradual transition would be based on foveal acuity as opposed to peripheral acuity, and we would expect that an ideal system would not require these transition zones.

A second explanation is that with a dynamically foveated system, eye-tracking error could create temporal artifacts at the boundary between zones similar to what we observed in frame-by-frame resampling in Experiment 1. Consider the case that the eye is fixated at the center of the display and a dynamically foveated zone with best acuity is rendered with a 10° radius about that central location, with downsampling beyond. If that 10° radius high-resolution zone was adjusted based on frame-by-frame eye tracking and the tracking was susceptible to $\pm 1^\circ$ of jitter, there would be a 2° border region around the high-resolution zone that flickered as it oscillated from being high resolution to low resolution. Applying a spatial transition zone that is at least 2° wide could greatly reduce the contrast of the temporal fluctuation and make any scintillation less objectionable. If tracking jitter were the explanation, one would expect that the

transition zone should be tuned based on the eye-tracker jitter and not on an attribute of peripheral vision. Alternatively, an appropriate temporal filter on the eye tracker signal could ensure that adjustments avoid certain scintillation frequencies that are objectionable. Exploring these mechanisms of foveated rendering failures at transition regions deserves extended study.

Related to eye tracking is another important question about foveated rendering: What should be the maximum latency of such a system. There are two major ways in which the latency could be relevant: with pursuit eye movements and with saccades. During a pursuit eye movement, a region in the periphery could be closer to the fovea than ideal by the product of the pursuit movement velocity and the latency. Smooth pursuit eye movements can attempt to track objects moving at speeds approaching $100^\circ/\text{s}$, although at this velocity, the smooth pursuit will be accompanied with catch up saccades.³⁶ Because of practical field of view limitations and pursuit lag, $10\text{--}20^\circ/\text{s}$ smooth pursuit is more typical.^{36,37} At $20^\circ/\text{s}$, even a comparatively slow 100 ms latency would only produce 2° of lag, which would be minor if the central full-resolution region were more than 10° radius. In the case of saccades, which can exceed $500^\circ/\text{s}$,³⁸ even a relatively small latency of a single frame could permit the fovea of the eye to land on a low-resolution portion of the image. The maximum time allowable before this low resolution can be replaced with a high-resolution rendering is of great interest for the development of the eye tracking and display pipeline.

There are several visual system attributes that could afford some tolerance for a longer latency loop. One is saccadic suppression, which has been reported to make the visual system insensitive to changes during and in the brief period following the saccade.³⁹ Another is temporal masking in which

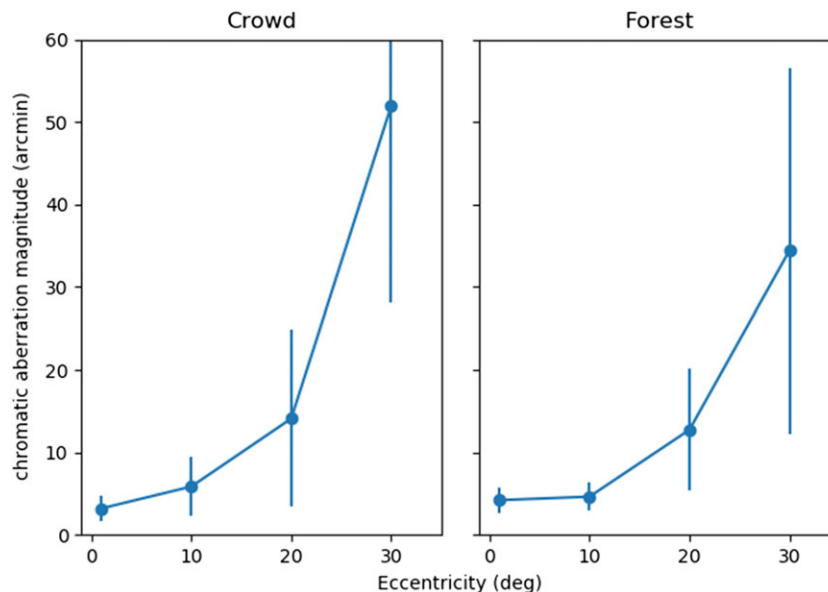


FIGURE 12 — Results of Experiment 3. The threshold displacement between the red and blue color planes is plotted as a function of the eccentricity.

the presentation of imagery immediately following a briefly presented stimulus prevents the visual system from perceiving the details.⁴⁰ In our experiment, we explored the latter effect. In the case of the zero eccentricity condition, we examine the case where the eye has fixated on a new location that is rendered with low resolution briefly before being re-rendered in high resolution. We also tested peripheral regions to see if they would be sensitive to brief presentations at low resolution before being replaced with high-resolution imagery. In these experiments, irrespective of eccentricity, we found that aggressively downsampled imagery became noticeable only if it was on screen for about 40 ms. It is possible that saccadic suppression could make this even more forgiving such as the 50–70 ms suggested by Albert and colleagues.³⁵

The use of molded plastic lenses to create a wide field of view has led to more chromatic dispersion than many other imaging systems. Understanding the sensitivity of the visual system to chromatic aberration can be valuable in determining how to best optimize the optics and the digital chromatic aberration correction. Our experiment showed that within 10° of the fovea, transverse chromatic aberration is detectable at the scale of arcminutes and decreases to about half a degree at 30° eccentricity. This follows a similar falloff as the temporally stable downsampling we observed in Experiment 1. This property of the visual system could enable a foveated chromatic aberration correction, in which chromatic aberration can go uncorrected in the periphery without being perceptible.

In the experiments, all testing was based on two images. The images were selected to be representative of content relevant for VR viewing. The differences in thresholds between the two images were fairly small. Conversely, based on the literature with more traditional acuity tasks, the difference in thresholds is highly task dependent, with performance falling off much more rapidly for hyperacuity and landolt C while only gradually falling off for letter recognition and identification.¹⁵ Furthermore, in these traditional acuity tasks, decreasing the stimulus contrast has a dramatic impact on the falloff with eccentricity. With continuous tone imagery, it is difficult to convert the specific image impairments into the primitives and contrast levels that have been tested in the literature. But based on the difference between the fovea and peripheral conditions, the sensitivity to the static artifacts (such as blur and stabilized aliasing) had falloff behavior that was relatively gradual and more similar to that of letter acuity than hyperacuity.

The work from these experiments can inform us about where we could best devote additional computation power both with current displays as well as future display systems. With today's systems, in the central 30° radius of the retina, the visual system is highly sensitive to temporal artifacts that are less than the size of a single 6 arcmin pixel. This suggests that anti-aliasing used in the rendering is valuable over a large portion of the retina. We could expect a brute force foveated rendering approach with today's pixel sizes to create visible artifacts if used in the central 60° cone. Beginning at a 20° radius, some foveated rendering is possible if techniques

are used to minimize the temporal volatility. Also, with the current pixel size, chromatic registration should ideally have less than a 3 arcmin offset in the central 20° but with increasing tolerance at 30° and beyond.

When the pixel size approaches the 1-arcmin target in which it is retinally limited at the fovea, aggressive downsampling will be possibly starting at only a few degrees and is limited primarily by the gaze tracking accuracy and latency.

Our results on temporal masking of low-resolution imagery with high-resolution imagery suggest that latency of about 40 ms would not be perceptible, regardless of the location in the retina. This finding is encouraging since if there were 60 Hz eye tracking, it could permit at least one frame for capturing the eye position post-saccade and one frame of the display system to be re-rendered and drawn to the display panel. The accuracy of the eye tracking is not something that we explicitly studied, but based on the results of the resolution boundary experiment, it seems likely that erratic gaze estimation could lead to a volatile resampling artifact in the periphery with sensitivity levels higher than a stable downsampling. Based on this hypothesis, gaze estimation instability could require a broader transition zone and possibly slightly expanded foveal zone.

Acknowledgments

We are grateful to the review and advice of our colleagues Haomiao Jiang, Peter Milford, and Nikhil Balram.

References

- 1 S. M. Anstis, "A chart demonstrating variations in acuity with retinal position," *Vision Res.* **31**, **14**, No. 7, 589–592 (1974).
- 2 B. Bastani *et al.*, "Foveated pipeline for AR/VR head-mounted displays," *Inform. Displays*, **33**, No. 6, 14–19 (2017).
- 3 C. Anthes *et al.*, State of the art of virtual reality technology. *Aerospace Conference*, 2016 *IEEE* (pp. 1–19), (2016).
- 4 C. J. Lin and B. H. Woldegiorgis, "Interaction and visual performance in stereoscopic displays: a review," *J. Soc. Inf. Disp.*, **23**, No. 7, 319–332 (2015).
- 5 P. Labatut *et al.*, Fast level set multi-view stereo on graphics hardware. *3D Data Processing, Visualization, and Transmission*, **3** (2006).
- 6 S. Seitz *et al.*, A comparison and evaluation of multi-view stereo reconstruction algorithms. *IEEE Computer Society Conference on Computer Vision and Pattern Recognition*, **1**, (2006).
- 7 B. Karis, High-quality temporal supersampling. *ACM SIGGRAPH Courses*, **10**, (2014).
- 8 K. Vaidyanathan *et al.*, Coarse pixel shading. *HPG '14 Proceedings of High Performance Graphics*, 9–18, (2014).
- 9 C. Wyman and M. McGuire, Hashed alpha testing. *I3D*, 1–7, (2017).
- 10 M. Weier *et al.*, "Foveated real-time ray tracing for head-mounted displays," *Eurographics*, **35** (2016).
- 11 Y. S. Pai *et al.*, "GazeSim: simulating foveated rendering using depth in eye gaze for VR," *SIGGRAPH* (2016).
- 12 B. Guenter *et al.*, "Foveated 3D graphics," *ACM Trans. Graph. (TOG)*, **36** (2012).
- 13 A. Patney *et al.*, "Towards foveated rendering for gaze-tracked virtual reality," *ACM Trans. Graph. (TOG)*, **35** (2016).
- 14 A. T. Duchowski and A. Coltekin, "Foveated gaze-contingent displays for peripheral LOD management, 3D visualization, and stereo imaging," *ACM Trans. Multimedia Comp., Comm. Appl. (TOMM)*, **24** (2007).

15 H. Strasburger *et al.*, "Peripheral vision and pattern recognition: a review," *J. Vision*, **11** (2011).

16 J. Lubin, A human visual system model for objective picture quality measurements. *International broadcasting conference (IBC)*, 498–503 (1997).

17 L. S. Stone *et al.*, "Linking eye movements and perception," *J. Vision*, **3** (2003).

18 D. M. Hoffman and D. Stolzka, "A new standard method of subjective assessment of barely visible image artifacts and a new public database," *J. Soc. Inf. Disp.*, **22**, No. 12, 631–643 (2014).

19 S. P. McKee and K. Nakayama, "The detection of motion in the peripheral visual field," *Vision Res.*, **24**, No. 1, 25–32 (1984).

20 E. Liu, "Lens matched shading and unreal engine 4 integration part 1," Online, published Jan 18, (2017). <https://developer.nvidia.com/lens-matched-shading-and-unreal-engine-4-integration-part-1>, Accessed 23-Jan-2018.

21 C. Vierl *et al.*, "An 18 megapixel 4.3" 1443 ppi 120 Hz OLED display for wide field of view high acuity head mounted displays," *J. Soc. Inform. Disp.*, **26**, No. 5, 314–324 (2018).

22 M. Fujita and T. Harada, "Foveated real-time ray tracing for virtual reality headset," *Light Transp. Entertainment Res.* (2014).

23 A. Vlachos. Advanced VR rendering performance. *Game Developers Conference*, (2016).

24 S. Hall, and J. Milner-Moore. Higher res without sacrificing quality and other lessons from playstation VR worlds. *Game Developers Conference*, (2017).

25 A. T. Bahill, "Most naturally occurring human saccades have magnitudes of 15 deg or less," *Invest. Ophthalmol.*, **14**, 468–469 (1975).

26 M. Haynes and T. Starner, "Effects of lateral eye displacement on comfort while reading from a video display terminal," *Proc. ACM on Interact., Mobile, Wearable Ubiquit. Technol.*, **1**, No. 4, 138 (2018).

27 W. Mason, "SMI showcases foveated rendering with new 250Hz eye tracking kit," Online, published Jan, (2016). <https://uploadvr.com/smi-eye-tracking-foveated-rendering-exclusive/>, Accessed 18-Oct-2017.

28 T. Lindeberg. Concealing rendering simplifications using gaze contingent depth of field. Master's Thesis, Royal institute of technology, school of computer science and communication, (2016).

29 M. Woo *et al.*, "OpenGL programming guide, 2nd edition," OpenGL Architecture Review Board, Addison-Wesley, 338–362, (1998).

30 H. Hoppe, Progressive meshes. In *Proceedings of the 23rd annual conference on Computer graphics and interactive techniques* (pp. 99–108). ACM, (1996).

31 E. Turner *et al.*, Phase-aligned foveated rendering for virtual reality headsets. *IEEE VR, the 25th IEEE Conference on Virtual Reality and 3D User Interfaces*, (2018).

32 M. Stengel *et al.*, "Adaptive image-space sampling for gaze-contingent real-time rendering," *Comp. Graph. Forum*, **35**, No. 4, 129–139 (2016).

33 L. O. Harvey Jr. and E. Pöppel, "Contrast sensitivity of the human retina," *Optom. Vis. Sci.*, **49**, No. 9, 748–753 (1972).

34 S. J. Anderson *et al.*, "Human peripheral spatial resolution for achromatic and chromatic stimuli: limits imposed by optical and retinal factors," *J. Physiol.*, **442**, No. 1, 47–64 (1991).

35 R. Albert *et al.*, *ACM Trans. Appl. Percept. (TAP)*, **14**, No. 4, 25 (2017).

36 C. H. Meyer *et al.*, "The upper limit of human smooth pursuit velocity," *Vision Res.*, **25**, No. 4, 561–563 (1985).

37 H. Collewijn and E. P. Tamminga, "Human smooth and saccadic eye movements during voluntary pursuit of different target motions on different backgrounds," *J. Physiol.*, **351**, No. 1, 217–250 (1984).

38 R. W. Baloh *et al.*, "Quantitative measurement of saccade amplitude, duration, and velocity," *Neurology*, **25**, No. 11, 1065–1065 (1975).

39 A. Thiele *et al.*, "Neural mechanisms of saccadic suppression," *Science*, **295**, No. 5564, 2460–2462 (2002).

40 J. T. Enns and V. Di Lollo, "What's new in visual masking?" *Trends Cogn. Sci.*, **4**, No. 9, 345–352 (2000).



David Hoffman graduated from the University of California San Diego with a degree in Bioengineering and received his PhD in Vision Science from the School of Optometry at the University of California Berkeley. He has since worked with several companies on improving displayed image quality through identifying, characterizing, and mitigating degradation and distortion sources throughout the software and hardware acquisition and display pipeline. Since 2017, he is an applied vision researcher at Google working on creating the display subsystem and imaging pipelines that facilitate terrific head worn display experiences. A major part of this research is discovering how the limits of the visual system dictate which technological advances will have the strongest perceptual impact. He is an associate editor of the *Journal of Society for Information Display* and chairs the applied vision subcommittee of SID Display Week.



Zoe Meraz graduated from the University of Nevada, Reno with a degree in Biochemistry and Molecular Biology in 2015. She has since worked in human factors and eye tracking at a startup in Milpitas and now focuses on visual human factors at Google Daydream.



Eric Turner is a software engineer at Google Daydream, focusing on foveated rendering techniques and 3D reconstruction. Eric received his BS in Electrical and Computer Engineering from Carnegie Mellon University in 2011, and his MS and PhD in Electrical Engineering and Computer Sciences from the University of California Berkeley in 2013 and 2015, respectively. His background is focused in 3D reconstruction and computational geometry, including working as the CTO and co-founder Indoor Reality, Inc. and working on rapid 3D modeling of building interiors.