# Supplementary Materials: Study Details
# Perceptual Rasterization for Head-mounted Display Image Synthesis

SEBASTIAN FRISTON, University College London
TOBIAS RITSCHEL, University College London
ANTHONY STEED, University College London

**ACM Reference format:**
Sebastian Friston, Tobias Ritschel, and Anthony Steed. 2019. Supplementary Materials: Study Details
Perceptual Rasterization for Head-mounted Display Image Synthesis. *ACM Trans. Graph.* 38, 4, Article 97 (January 2019), 6 pages.
DOI: 10.1145/3306346.3323033

## 1 USER STUDY

This supplemental document provides details on the perceptual experiments from Section 6 in the main article.

### 1.1 Apparatus

Figure 1 shows the eye tracking apparatus used for the Foveated Rendering user studies (1.2 & 1.4). We used a 2018-typical desktop PC with a Nvidia GTX 980 GPU and an Asus VG248 144 Hz monitor to render the stimuli. The eye-tracker was an SR-Research EyeLink II connected via Ethernet directly to our application. Study 3 (1.5) (Object Tracking) used a standard Windows desktop driving an Oculus DK2.

(a) Eye-Link II Headset (b) Complete Apparatus

Fig. 1. The apparatus used in user studies 1 & 2

### 1.2 Study 1: Fovetion strength

This study demonstrated that there is nearly no perceptual difference between a non-trivially foveated image and a traditionally rendered image. To this end, we fit a function relating the level of foveation and the probabilityof detecting it. Such a psychmetric function can predict that, e. g., 75 % of all subjects would not note a difference for a certain foveation level.


DOI: 10.1145/3306346.3323033

This protocol follows Patney et al. [2], whor performed a 2AFC staircase task to identify the Just Noticeable Difference (JND) threshold - the foveation strength at which participants begin to reliably detect foveation artefacts.

*1.2.1 Procedure.* After being fitted with the eye-tracker (Section 1.1), participants engaged in a series of Two-Alternative Forced Choice (2AFC) trials. In each trial, participants were exposed to two 1.5 second sequences of a rotating model - one with foveation and one traditionally rendered - with a .75-second gap in-between. The rotation was around the vertical axis at one revolution every 14 seconds. After viewing both sequences, participants were asked to indicate via the keyboard which of the two was "higher quality". The order of rendering technique was randomized. Participants each completed 180 trials on one of three models. Foveation strength was determined by a 1-up/3-down staircase following the guidelines of Garcia-Perez & Alcala-Quintana [1].

*1.2.2 Stimuli.* Figure 2 shows the stimuli for User Study 1. These screen captures were taken with a foveation strength of 0.01 (the starting strength). Figure 3 gives examples of these stimuli with high foveation (10) strengths and an off-center (0.9,0.9) foveal viewpoint. Note that this strength is far higher than any users tolerance, it is increased here to ensure the effect is visible in scaled down images.

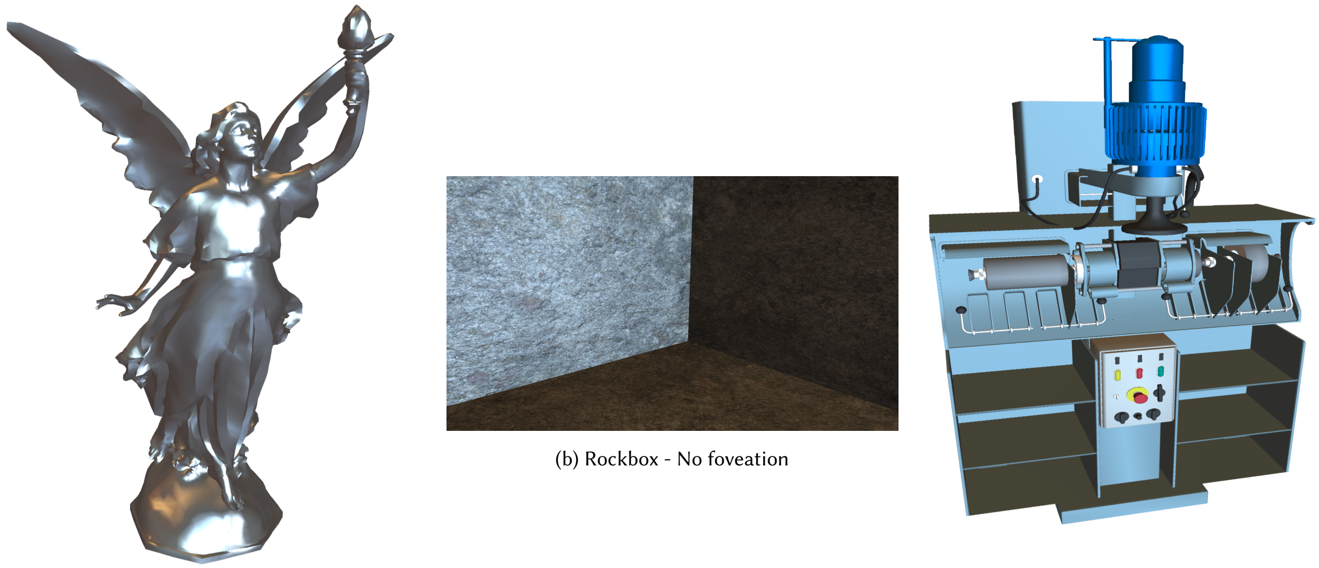
*1.2.3 Participants.* 25 naïve participants successfully completed our study across three conditions: Lucy (7), Rockbox (9), CAD (9). This study was approved under UCL REC 5998 006.

*1.2.4 Analysis.* We opted for a fixed-size staircase with empirically set step-sizes, as our technique is novel and we do not have any reasonable priors for parametric sampling schemes. For our analysis though we fit a logistic psychometric function [1] for simplicity and comparability using `Psignifit 4` [3], to estimate thresholds and confidence intervals at 95 %.

*1.2.5 Staircases.* The foveation strength user study used a 1-up/3-down staircase with an up/down ratio of .7393. Figure 4 shows the staircases for individual participants for each condition (model).

*1.2.6 Approximate Thresholds.* An approximate threshold can be computed by averaging the reversals in a staircase. The individual thresholds for each participant are shown in Table 1.

*1.2.7 Psychometric Functions.* Averaging reversals provides a rough approximation. To better understand the nature of the participants perception of the artefacts we fit psychometric functions. A psychometric function describes the probability of detecting a distortion (vertical axis) depending on the foveation strength (horizontal). The functions for each participant, for each condition, are shown

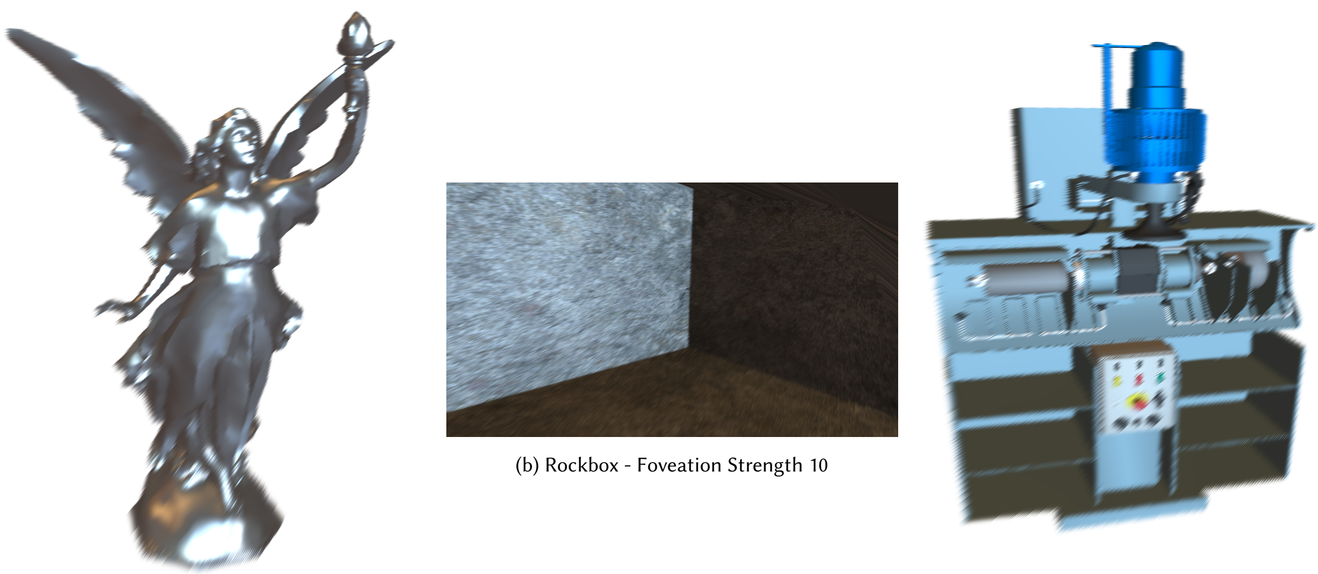


(a) Lucy - No foveation

(b) Rockbox - No foveation

(c) CAD - No foveation

Fig. 2. Example stimuli from User Study 1 for each condition

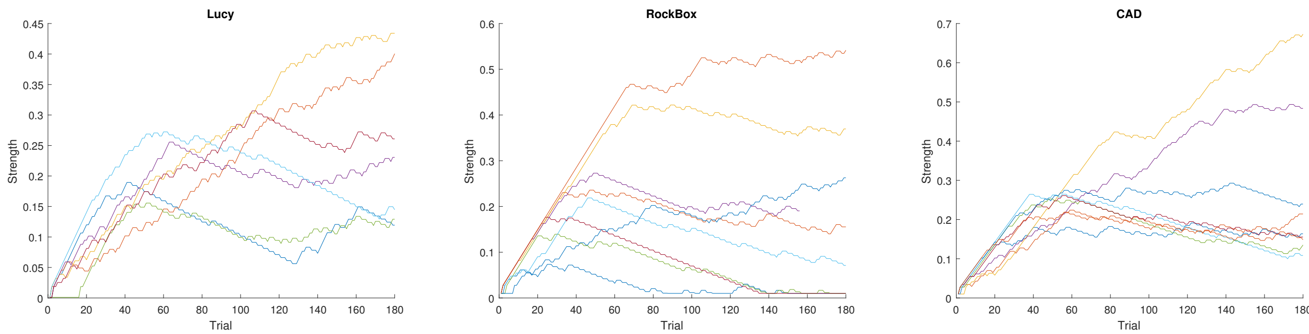


(a) Lucy - Foveation Strength 10

(b) Rockbox - Foveation Strength 10

(c) CAD - Foveation Strength 10

Fig. 3. Example Foveation Strength stimuli with a high (exaggerated for print reproduction) level of foveation.




(a) Lucy (b) Rockbox (c) CAD

Fig. 4. Per-participant staircases for each Condition in User Study 1

Table 1. Threshold Estimates each participant, for each condition, in User Study 1

| | Condition | | |
|---|---|---|---|
| Participant ID | Lucy | Rockbox | CAD |
| 1 | 0.1137 | 0.0304 | 0.1590 |
| 2 | 0.2177 | 0.1843 | 0.1546 |
| 3 | 0.2913 | 0.3723 | 0.4043 |
| 4 | 0.1715 | 0.1982 | 0.3488 |
| 5 | 0.1110 | 0.0669 | 0.1693 |
| 6 | 0.2207 | 0.1107 | 0.1986 |
| 7 | 0.1860 | 0.1448 | 0.1824 |
| 8 | | 0.1609 | 0.2505 |
| 9 | | 0.4844 | 0.1755 |

in Figure 5. We see that the 75 % detection probability JND threshold occurs at non-zero levels of foveation. This indicates subjects cannot detect our foveation even when present at such strengths. The confidence intervals (colored bars) show the significance of the observation.

The parameter estimates from the functions are shown for each participant in Table 2. The lower asymptote (guess rate) is always 0.5 due to the chosen protocol. *Eta* ([0, 1]) is a measure of dispersion with zero indicating a low rate of dispersion. The documentation for `psignifit 4` is available at https://github.com/wichmann-lab/psignifit/wiki/Basic-Usage.

## 1.3 Further Analysis

The average thresholds are $\overline{\text{Lucy}}$: 0.23, $\overline{\text{Rockbox}}$: 0.16 & $\overline{\text{CAD}}$: 0.32. The only condition to have thresholds potentially at zero was RockBox. This could be because participants detected artefacts with baseline foveation, but due to the nature of the staircases (no participant detected a difference until a non-trivial threshold was reached) and infrequency, it is also possible they unintentionally self-trained on a secondary cue. The thresholds encapsulate many factors, including eye tracker calibration, latency and screen resolution, so the exact value is specific to our apparatus.

The important conclusion however is that they are non-zero, demonstrating our technique does work. If it were integrated into a more tightly controlled system such as an HMD, we would expect thresholds to increase.

## 1.4 Study 2: Foveation preference

We performed a second study to determine how users compare our foveally super-sampled images to traditionally rendered images, when presented with a super-sampled image as a reference.

*1.4.1 Procedure.* After being fitted with the eye-tracker (Section 1.1), participants engaged in a series of 2AFC trials. In each trial, participants viewed three instances of a slowly rotating model side-by-side. The center model was a 4×4 super-sampled reference, with common and foveated to the sides in a randomized order. Participants were asked to indicate via the keyboard which side appeared most similar to the reference. Subjects each completed 45 trials spread evenly across three conditions, randomly interleaved.

*1.4.2 Stimuli.* Figure 6 shows the stimuli for User Study 2.

*1.4.3 Participants.* 7 naïve participants completed this study. This study was approved under UCL REC 5998 006.

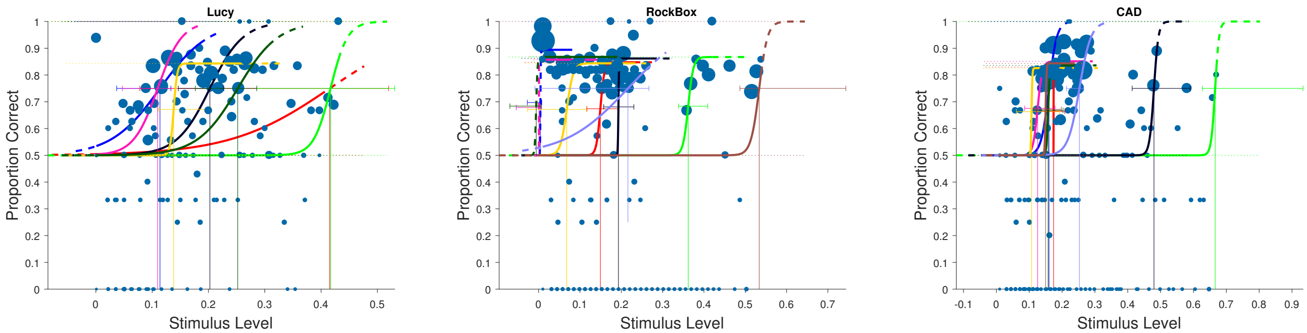
*1.4.4 Analysis.* We compute the preference for foveation as the proportion of aggregated trials in which the foveated image was chosen, for each condition. Two-tailed binomial tests ($n = 105$) indicated that the preferences are significantly different to chance ($p = .5$) for all conditions ($p < .001$);

*1.4.5 Results.* Figure 7 shows the individual proportions for each participant, for each model in User Study 2. The proportions in all cases are out of 15 trials. The same results are listed in Table 3.

The results show a strong and consistent preference for foveation across all models ($\overline{\text{Lucy}}$: 90%, $\overline{\text{Flower}}$: 94%, $\overline{\text{CAD}}$: 87%). The slight reduction for CAD is likely because when viewing the back of the model there were few details to distinguish the techniques.

## 1.5 User Study 3: Object Tracking

We conducted a user study to examine how rolling rasterization affects perception in Virtual Reality (VR). We used an object tracking task to measure how behavior is influenced by both rolling




(a) Lucy (b) Rockbox (c) CAD

Fig. 5. Psychometric Functions fitted for each participant, for each condition in User Study 1

Table 2. Psychometric Function Parameters for each participant in condition Lucy

| | Participant | Threshold | | | Width | | | Lapse Rate | | | Eta | | |
|---|---|---|---|---|---|---|---|---|---|---|---|---|---|
| | | Estimate | Lower | Upper | Estimate | Lower | Upper | Estimate | Lower | Upper | Estimate | Lower | Upper |
| Lucy | 1 | 0.114 | 0.037 | 0.133 | 0.256 | 0.056 | 0.435 | 0.000 | 0.004 | 0.211 | 0.000 | 0.002 | 0.179 |
| | 2 | 0.415 | 0.078 | 0.520 | 0.520 | 0.114 | 0.990 | 0.000 | 0.004 | 0.373 | 0.000 | 0.002 | 0.230 |
| | 3 | 0.417 | 0.286 | 0.531 | 0.089 | 0.055 | 0.966 | 0.000 | 0.003 | 0.289 | 0.000 | 0.002 | 0.193 |
| | 4 | 0.203 | 0.147 | 0.227 | 0.173 | 0.031 | 0.445 | 0.000 | 0.003 | 0.219 | 0.000 | 0.002 | 0.213 |
| | 5 | 0.110 | 0.047 | 0.141 | 0.126 | 0.022 | 0.352 | 0.000 | 0.002 | 0.214 | 0.000 | 0.009 | 0.434 |
| | 6 | 0.138 | 0.111 | 0.205 | 0.019 | 0.016 | 0.459 | 0.157 | 0.009 | 0.214 | 0.000 | 0.002 | 0.193 |
| | 7 | 0.252 | 0.177 | 0.286 | 0.211 | 0.060 | 0.588 | 0.000 | 0.003 | 0.218 | 0.000 | 0.002 | 0.184 |
| Rock Box | 1 | 0.004 | -0.027 | 0.008 | 0.003 | 0.004 | 0.146 | 0.106 | 0.052 | 0.179 | 0.000 | 0.003 | 0.267 |
| | 2 | 0.149 | 0.117 | 0.194 | 0.029 | 0.014 | 0.380 | 0.152 | 0.008 | 0.221 | 0.000 | 0.003 | 0.277 |
| | 3 | 0.362 | 0.339 | 0.409 | 0.037 | 0.021 | 0.328 | 0.133 | 0.003 | 0.209 | 0.000 | 0.003 | 0.413 |
| | 4 | 0.193 | 0.157 | 0.230 | 0.004 | 0.013 | 0.362 | 0.138 | 0.007 | 0.220 | 0.000 | 0.003 | 0.291 |
| | 5 | 0.001 | -0.054 | 0.007 | 0.000 | 0.008 | 0.302 | 0.143 | 0.076 | 0.208 | 0.000 | 0.002 | 0.222 |
| | 6 | 0.068 | -0.027 | 0.115 | 0.046 | 0.022 | 0.530 | 0.156 | 0.011 | 0.217 | 0.000 | 0.002 | 0.165 |
| | 7 | -0.006 | -0.070 | 0.006 | 0.010 | 0.009 | 0.374 | 0.132 | 0.078 | 0.213 | 0.000 | 0.006 | 0.311 |
| | 8 | 0.216 | -0.009 | 0.267 | 0.452 | 0.077 | 0.670 | 0.000 | 0.005 | 0.311 | 0.000 | 0.002 | 0.240 |
| | 9 | 0.534 | 0.487 | 0.744 | 0.059 | 0.031 | 0.760 | 0.000 | 0.003 | 0.357 | 0.508 | 0.210 | 0.742 |
| CAD | 1 | 0.160 | 0.127 | 0.175 | 0.070 | 0.014 | 0.257 | 0.000 | 0.002 | 0.187 | 0.000 | 0.003 | 0.313 |
| | 2 | 0.174 | 0.127 | 0.208 | 0.001 | 0.023 | 0.437 | 0.171 | 0.003 | 0.222 | 0.000 | 0.002 | 0.229 |
| | 3 | 0.666 | 0.627 | 0.934 | 0.036 | 0.037 | 1.018 | 0.000 | 0.003 | 0.381 | 0.000 | 0.006 | 0.401 |
| | 4 | 0.480 | 0.413 | 0.590 | 0.040 | 0.025 | 0.856 | 0.000 | 0.002 | 0.252 | 0.000 | 0.002 | 0.194 |
| | 5 | 0.126 | 0.086 | 0.178 | 0.042 | 0.015 | 0.491 | 0.149 | 0.011 | 0.221 | 0.000 | 0.002 | 0.207 |
| | 6 | 0.107 | 0.061 | 0.193 | 0.007 | 0.017 | 0.588 | 0.174 | 0.008 | 0.232 | 0.001 | 0.002 | 0.187 |
| | 7 | 0.158 | 0.123 | 0.203 | 0.015 | 0.015 | 0.431 | 0.164 | 0.008 | 0.225 | 0.000 | 0.002 | 0.205 |
| | 8 | 0.253 | 0.214 | 0.266 | 0.103 | 0.027 | 0.272 | 0.000 | 0.003 | 0.184 | 0.000 | 0.002 | 0.252 |
| | 9 | 0.150 | 0.116 | 0.198 | 0.008 | 0.015 | 0.364 | 0.156 | 0.004 | 0.220 | 0.000 | 0.004 | 0.364 |

rasterization and the asynchronous time-warping. This study was approved under UCL REC 5998 003.

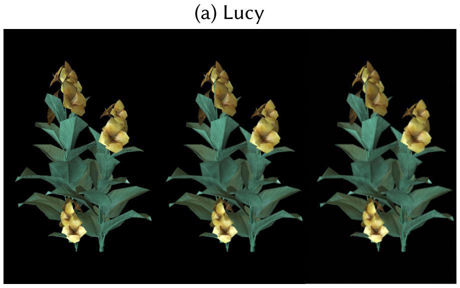
*1.5.1 Apparatus.* This experiment was performed with an Oculus Rift DK2. This HMD has a single low-persistence rolling-scanout display that scans right-to-left at 75 Hz with a persistence of 4 ms. The DK2 has an inertial measurement unit (IMU) that samples at 1 kHz. The head and box positions were sampled at 75 Hz.

*1.5.2 Procedure.* Participants were shown a simple virtual environment in which a 50 cm box moved along a 180° curve, 8 m in front of them, just above eye level. The box reversed direction at the extents and moved at $85.9 \pm 68.7\,^\circ\,s^{-1}$, the rate changing randomly every second. A head-fixed reticle was visible 8 m ahead. These parameters were set by pre-trials.

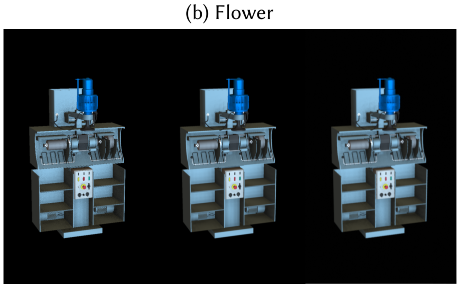
Participants were told to use their head to keep the reticle in the middle of the box. Participants followed the box in two trials, each lasting 4 minutes. Between the trials participants had a short break outside the Head-Mounted Display (HMD). The task was difficult - participants rarely, if ever, acquired the box. There were three conditions pertaining to the rasterization method used: traditional (STD), Oculus' Asynchronous Time-warping (ATW) and our Rolling-Rasterization (ROL). The conditions were presented in 30-second-blocks, randomly interleaved.

*1.5.3 Stimuli.* Figure 9 shows the stimuli that participants were following.

(a) Lucy



(b) Flower



(c) CAD

Fig. 6. Example stimuli from User Study 2 for each condition

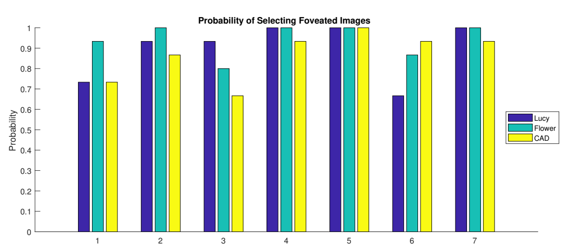



Fig. 7. Probability of selecting Foveated images as most similar to Reference across 15 trials

*1.5.4 Participants.* 20 naïve participants completed the study. All started both trials. One aborted the second trial mid-way due to simulator sickness. This study was approved under UCL REC 5998 003.

*1.5.5 Analysis and Results.* We began by analyzing the phase of the head motion. This is the instantaneous angular difference between the head and the box, positive when the head is leading the box, and negative when it is following. The phase probability distribution is shown in Figure 8. If participants were tracking the box exactly, we would expect a symmetrical distribution with a slight negative bias due to latency. Instead, a set of Kolmogorov-Smirnov tests show a non-normal distribution for the conditions separately and cumulatively ($P < 0.05$). All conditions also show a positive bias. This indicates a tendency to lead the target.

In this case, we would expect the lead to increase as apparent latency decreases, and this is what is shown. ROL enables the largest anticipatory behavior (a lead of $\overline{2.7}^\circ$). STD presents the second largest ($\overline{2.2}^\circ$), having a latent but head-fixed image. ATW has the smallest lead ($\overline{1.7}^\circ$), because while it compensates for latency, it does so by moving the entire image - including the target - counter to head rotation introducing apparent lag into the target.

To test the significance of this, we performed an ANOVA on the terms of a linear-mixed model, to control for per-participant biases and speed. The results of this test are shown in Table 4, indicating a highly significant effect of rendering condition ($p < .0001$), as well as an interaction between rendering condition and speed ($p < .0001$).

## REFERENCES

[1] Miguel García-Pérez and Rocío Alcalá-Quintana. 2007. A comparison of fixed-step-size and Bayesian staircases for sensory threshold estimation. *Spatial Vis.* 20, 3 (2007), 197–218.

[2] Anjul Patney, Marco Salvi, Joohwan Kim, Anton Kaplanyan, Chris Wyman, Nir Benty, David Luebke, and Aaron Lefohn. 2016. Towards foveated rendering for gaze-tracked virtual reality. *ACM Trans. Graph. (Proc. SIGGRAPH)* 35, 6 (2016), 179.

Table 3. Probability of selecting Foveated images as the most similar to Reference across 15 trials

| Participant | Probability | | |
|---|---|---|---|
| | Lucy | Flower | CAD |
| 1 | 0.73 | 0.93 | 0.73 |
| 2 | 0.93 | 1.00 | 0.87 |
| 3 | 0.93 | 0.80 | 0.67 |
| 4 | 1.00 | 1.00 | 0.93 |
| 5 | 1.00 | 1.00 | 1.00 |
| 6 | 0.67 | 0.87 | 0.93 |
| 7 | 1.00 | 1.00 | 0.93 |

Table 4. ANOVA test results for mixed-model terms

| Variable | Effect | F-Stat | DoF | DoF | $p$ |
|---|---|---|---|---|---|
| Per-Parti. Inter. | Random | 1394.14 | 1 | 501,799 | <.0001 |
| Condition | Fixed | 114.53 | 2 | 501,799 | <.0001 |
| Speed | Fixed | 7984.79 | 1 | 501,799 | <.0001 |
| Condition: Speed | - | 134.06 | 2 | 501,799 | <.0001 |

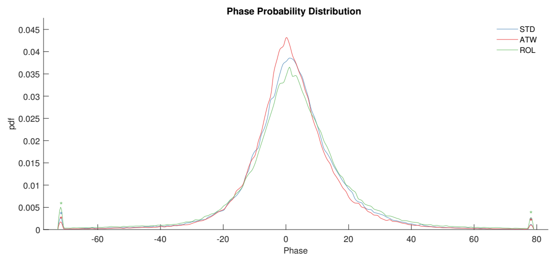



Fig. 8. Aggregate Phase Distributions from User Study 3 for each Condition

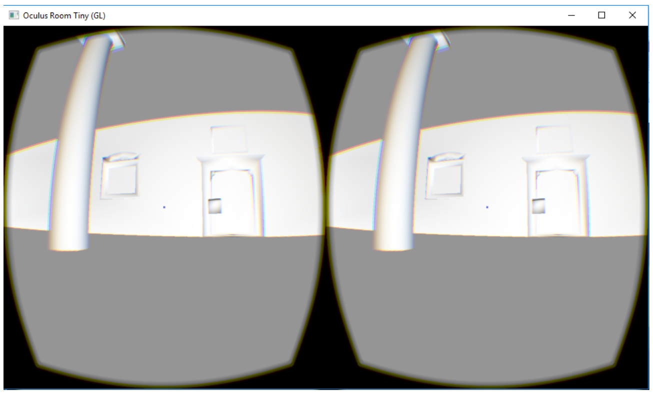



Fig. 9. Environment containing moving box observed by participants during the Tracking task

[3] Heiko H. Schütt, Stefan Harmeling, Jakob H. Macke, and Felix A. Wichmann. 2016. Painfree and accurate Bayesian estimation of psychometric functions for (potentially) overdispersed data. *Vis. Res.* 122 (2016), 105–123.