

Towards Attention-aware Foveated Rendering

BROOKE KRAJANCICH, Stanford University, USA

PETR KELLNHOFER, TU Delft, Netherlands

GORDON WETZSTEIN, Stanford University, USA

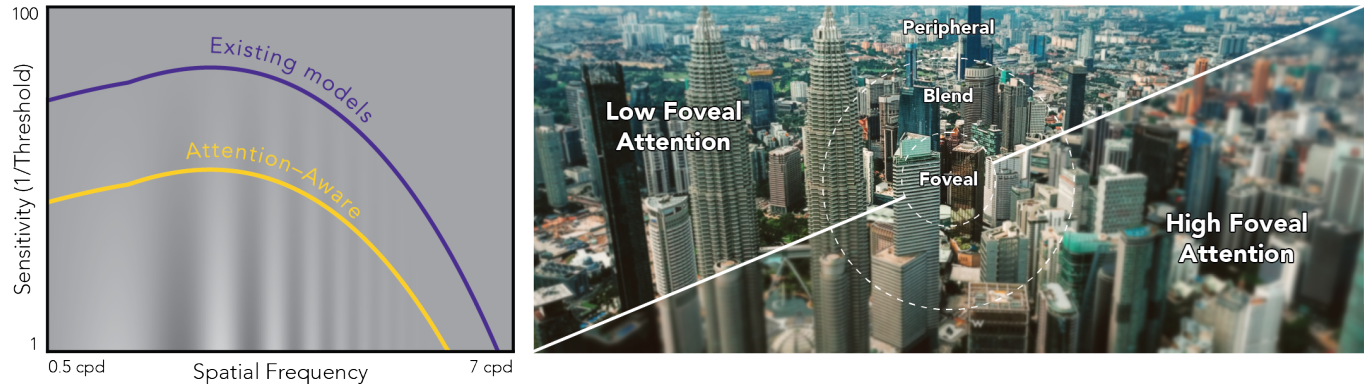


Fig. 1. Foveated graphics techniques rely on eccentricity-dependent models of human vision. However, existing models of contrast sensitivity (left, purple line, shown for a fixed eccentricity) do not take into account allocation of visual attention across the visual field. Our work is the first to experimentally derive a model for eccentricity-dependent attention-aware sensitivity (left, yellow line). As illustrated on the right, when the user is focused on a task in the fovea, less attention is directed to the periphery and a higher level of foveation (i.e., peripheral blur) is possible without impacting the perceived visual quality.

Foveated graphics is a promising approach to solving the bandwidth challenges of immersive virtual and augmented reality displays by exploiting the falloff in spatial acuity in the periphery of the visual field. However, the perceptual models used in these applications neglect the effects of higher-level cognitive processing, namely the allocation of visual attention, and are thus overestimating sensitivity in the periphery in many scenarios. Here, we introduce the first attention-aware model of contrast sensitivity. We conduct user studies to measure contrast sensitivity under different attention distributions and show that sensitivity in the periphery drops significantly when the user is required to allocate attention to the fovea. We motivate the development of future foveation models with another user study and demonstrate that tolerance for foveation in the periphery is significantly higher when the user is concentrating on a task in the fovea. Analysis of our model predicts significant bandwidth savings over those afforded by current models. As such, our work forms the foundation for attention-aware foveated graphics techniques.

CCS Concepts: • **Hardware** → **Displays and imagers**; • **Computing methodologies** → **Computer graphics**; **Mixed / augmented reality**.

Additional Key Words and Phrases: applied perception, visual attention, foveated rendering, foveated compression, virtual reality, augmented reality

1 INTRODUCTION

Virtual and augmented reality (VR/AR) are next-generation display systems that promise perceptually realistic user experiences by matching the resolution of human vision across a wide field of view. However, the necessary bandwidth for rendering, streaming, and displaying the required visual data is not yet possible with current technology. Foveated graphics has emerged as a suite of techniques

that exploits the eccentricity dependent acuity of human vision to minimize bandwidth in an imperceptible manner. In VR/AR, this is often implemented using gaze-contingent rendering, shading, compression or display (see Sec. 2.1). While these methods build on the insight that the human visual system has a limited ability to sense spatio-temporal changes in light, they have yet to consider how this might be dependent on higher-level cognitive processing.

Indeed, research shows that we rarely see what we are looking at unless we direct sufficient cognitive resources [Mack 2003]. This explains many phenomena, including change blindness or tunnel vision. *Visual attention* refers to a set of cognitive operations that helps us selectively process the vast amounts of information with which we are confronted, allowing us to focus on a certain location or aspect of the visual scene, while ignoring others. Most often, we direct our attention *overtly*, by moving our eyes towards a location, but we can also direct attention to an area in the periphery *covertly*, via a mental shift. Several studies have demonstrated that, under many conditions, increasing the amount of attention allocated to a visual task can enhance performance [Lee et al. 1997; Sperling and Melchner 1978]. In a similar manner, dividing attention between tasks reduces contrast sensitivity [Huang and Dobkins 2005; Mahjoob and Anderson 2019], visual acuity [Mahjoob et al. 2022], and speed of information accrual [Carrasco et al. 2006].

However, existing models of contrast sensitivity and visual acuity across the visual field are built on experiments where subjects are asked to covertly direct high levels of visual attention to a discrimination task in the periphery. Thus, for most scenarios in the real world and VR/AR, where most of our attention is directed overtly (at

Authors' addresses: Brooke Krajancich, Stanford University, USA, brookek@stanford.edu; Petr Kellnhofer, TU Delft, Netherlands, p.kellnhofer@tudelft.nl; Gordon Wetzstein, Stanford University, USA, gordon.wetzstein@stanford.edu.

our gaze position), we are likely overestimating our perceptual abilities in the periphery. Consequently, current efficacies of foveated graphics are too conservative in most real-use cases.

In this paper, we propose to account for the effect of covert attention when modeling human contrast sensitivity. To this goal, we investigate the effect of modulating the amount of attention allocated to the contrast discrimination task in the periphery, by forcing attention to the fovea with a visually demanding task. Specifically, we compare the standard approach to measuring contrast sensitivity, where a low amount of attention is directed to the fovea (“low”), to scenarios where part or most of the attention is directed there (“medium” and “high”). We show that in such instances, peripheral contrast discrimination thresholds elevate significantly, introduce the first attention-aware contrast sensitivity model, and motivate the development of future foveation models that take this into account.

To summarize, we make the following contributions:

- We design and conduct user studies to measure and validate eccentricity-dependent effects of attention on contrast discrimination and foveation efficacy.
- We introduce the first analytic model of contrast sensitivity across eccentricity under varying attention.
- We analyze bandwidth considerations and demonstrate that our model may afford significant bandwidth savings over existing foveated graphics techniques.

Overview of Limitations. The primary goals of this work are to develop the first perceptual model for attention-dependent contrast sensitivity and to demonstrate its potential benefits to foveated graphics. However, we do not attempt to derive a measurement instrument for attention allocation across the visual field, nor do we propose new foveated rendering algorithms or specific compression schemes that directly use this model.

2 RELATED WORK

2.1 Foveated Graphics

Foveated graphics techniques exploit eccentricity-dependent aspects of human vision, such as acuity, to minimize the bandwidth of a graphics system by optimizing bit depth [McCarthy et al. 2004], color-fidelity [Duinkharjav et al. 2022], level-of-detail [Luebke and Hallen 2001; Murphy and Duchowski 2001], or by simply reducing the number of vertices or fragments a graphics processing unit has to sample, ray-trace, shade, or transmit to the display; see [Duchowski et al. 2004; Koulieris et al. 2019] for a review of this area. Foveated rendering is perhaps the most well-known example of this class of algorithms [Deng et al. 2022; Friston et al. 2019; Geisler and Perry 1998; Guenter et al. 2012; Kaplanyan et al. 2019; Patney et al. 2016; Sun et al. 2017; Tariq et al. 2022; Tursun et al. 2019]. The perceptual models underlying foveated graphics usually exploit spatial aspects of eccentricity-dependent human vision but, to the best of our knowledge, none of them model cognitive or attentional effects of human vision, which we aim to address with our work.

2.2 Eccentricity-dependent CSF Models

The human visual system (HVS) is limited in its ability to sense variations in light intensity over space and time. Such visual performance is often described by the spatio-temporal contrast sensitivity function (CSF), defined as the inverse of the contrast discrimination threshold, that is, the smallest contrast of sinusoidal grating that can be perceived at each spatial and temporal frequency [Robson 1966]. The CSF can be used to describe the gamut of visible spatio-temporal signals as well as the decrease in relative sensitivity with retinal eccentricity.

While the CSF has been studied for over 70 years [Robson 1966], Kelly [1979] was the first to fit an analytical function, although limited to the fovea and a single luminance. Similarly, Watson et al. [2016] devised the pyramid of visibility, a simplified model that can be used if only higher frequencies are relevant. This model also captured luminance dependence and was later refit to model stationary content at higher eccentricities [Watson 2018]. Recently, models capturing eccentricity dependence for the full spatio-temporal domain were also presented [Krajancich et al. 2021; Mantiuk et al. 2021; Tursun and Didyk 2022]. Notably, Mantiuk et al. [2022] proposed a unified model, StelaCSF, which accounts for all major dimensions of the stimulus: spatial and temporal frequency, eccentricity, luminance, and area by combining data from several previous papers. Similar to luminance contrast, sensitivity to color contrast has also been studied and a spatio-chromatic CSF for foveal vision has recently been fitted [Mantiuk et al. 2020].

However, the data used to fit eccentricity-dependent models is collected from experiments where subjects covertly direct high levels of visual attention to the contrast discrimination task in the periphery. Such a scenario is unlikely to be representative of viewing conditions in the real world or in VR/AR settings, and may overestimate perceptual thresholds.

A related perceptual quality relevant for foveation is visual acuity defined as the smallest resolvable image detail. While our work focuses on measurements of the CSF alone, we expect similar effects to apply to acuity as well, because acuity can be understood as the outer limit of the CSF gamut where the sensitivity of vision drops sharply.

2.3 Attention-dependent CSF

Visual attention lies at the crossroads between perception and cognition, allowing us to select relevant sensory information for preferential processing. It is often modeled as a “zoom” or “variable-power lens”, that is, the attended region can be adjusted in size, but defines a tradeoff between its size and processing efficiency because of limited processing capacities [Eriksen and St. James 1986]. Physiologically, attention modulates neuronal responses and alters the profile and position of receptive fields near the attended location [Anton-Erxleben and Carrasco 2013]. Behaviorally, it improves performance in various visual tasks. One prominent effect of attention is the modulation of performance in tasks that involve the visual system’s spatial resolving capacity [Carrasco 2018].

In line with the “zoom lens” model, several studies have shown that covert attention enhances contrast sensitivity at the attended location at the cost of decreased sensitivity at unattended locations

across the visual field, at different eccentricities and isoeccentric (polar angle) locations [Cameron et al. 2002; Carrasco 2011; Carrasco et al. 2000; Mahjoob and Anderson 2019]. In such studies, attention is typically modulated by visual means, e.g., pre-cueing the location of the visual target [Cameron et al. 2002; Carrasco et al. 2000], or by drawing attention away from the stimuli by use of a concurrent visual task presented elsewhere [2005; 2004]. Carrasco et al. [2000] found that pre-cueing attention to the visual target enhanced contrast sensitivity between 0.05 and 0.1 log units over a broad range of spatial frequencies, and later, Carrasco et al. [2006] described this attention effect as equivalent to applying an effective *contrast gain* to the stimulus. A similar effect also occurs for visual acuity [Montagna et al. 2009] and speed of information accrual [Carrasco et al. 2006].

Also in line with the “zoom lens” model, and most similar to our work, Huang and Dobkins [2005] showed that when attention is divided across several points in the visual field, this reduces its enhancement effect at each location. In particular, they showed that drawing attention to the fovea with a rapid serial visual presentation task reduced contrast discrimination performance in the periphery by up to a factor of 10.

However, each of these studies measures a single position in eccentricity in the perifovea ($5 - 10^\circ$) and often only for 1 or 2 users. To the best of our knowledge, this effect has not been modeled over the visual field, nor is there any available data for the effect in the periphery ($> 10^\circ$), which we hope to rectify with our work.

3 A MODEL FOR PERCEPTION UNDER DIVIDED ATTENTION

While modulating the amount of attention has been shown to affect contrast discrimination thresholds (see Sec. 2.3), insufficient data and a lack of existing models prevent this effect from being applied to existing CSF models and hence foveated graphics. In this section, we provide a detailed discussion of the user study we conducted and the model we fit to predict the effect of modulating peripheral attention on the CSF.

3.1 Measuring CSF

The CSF model we wish to acquire could be parameterized by temporal frequency, spatial frequency, rotation angle, eccentricity (i.e., distance from the fovea), direction from the fovea (i.e., temporal, nasal, etc.) and other parameters. However, due to the fact that each data point needs to be recorded for multiple subjects and for many contrasts per subject to determine the respective CSFs, sampling all dimensions at once seems infeasible. Instead, we nominally select 3 points across the eccentricity (e) available with our display (see Sec. 3.3), a spatial frequency (f_s) of 2 cpd, and a diameter of 5° for the furthest point. We then use the *cortical magnification factor* to scale the spatial frequencies and diameters at the other retinal positions (see stimuli No. 1–3 in Table 1) such that the discrimination thresholds should be approximately the same (see Supplement for more detail).

In order to obtain the contrast discrimination thresholds, we use Gabor patches, i.e., sinusoidal gratings modulated by a Gaussian function (see Supplement for more detail), as is used by most previous works (Table 1, [Mantiuk et al. 2022]). During each trial, 2

Table 1. Parameters of tested Gabor patches. For measuring the model (shown above the divider), we chose a diameter of 5° at the highest eccentricity of 21° to utilize the full field of view of our display and a spatial frequency f_s of 2 cpd, then use the cortical magnification factor to scale these parameters at 7° and 14° eccentricity. Gabor’s sigma was defined as 20% of the diameter. For validation, we chose 2 sets of Gabor parameters (shown below the divider) used to fit StelaCSF [2022], namely measurements taken by Virsu and Rovamo [1979] and Wright and Johnson [1983]. Stimulus No. 5 was also tested at 2 additional adaption luminances, 58 and 116 cd/m^2 (No. 6 and 7).

No.	Eccentricity ($^\circ$)	Diameter ($^\circ$)	Spatial Freq. (cpd)	Adaptation Lum. (cd/m^2)
1	7	2.16	4.62	28
2	14	3.58	2.79	28
3	21	5	2	28
4	9.25	1.7	2	28
5	15	5	4	28
6	15	5	4	58
7	15	5	4	116

patches are simultaneously presented for 500 ms, centered at a given eccentricity, to the left and right side of the central fixation position. Each grating is randomly orientated either horizontally or vertically and the user is asked to discriminate whether the patches are of the same or different orientations.

3.2 The Attention-modulating Task

Inspired by Huang and Dobkins [2005], we present a rapid serial visual presentation (RSVP) at the fixation cross in order to modulate the amount of attention paid to the peripheral contrast discrimination task. The RSVP stimulus consists of N $1^\circ \times 1^\circ$ letters, each lasting $500/N$ ms with 0 ms blank in between, such that the task lasts the total display duration of the peripheral Gabor patches. The color of the letters alternate between red and green (scaled to be approximately isoluminant with the background), where the initial color is randomized across trials, and the user is asked to identify the color of the “target letter” (the letter “T”, which appears only once in a given sequence). Increasing N increases the difficulty of the task and should force more attention to the fovea, at the cost of reduced attention to the periphery. Consequently, three task levels were chosen to have an N of 1 (easy), 4 (medium) and 6 (hard), to force “low”, “medium”, and “high” levels of attention to the fovea. The target letter “T” was also adjusted such that for the “medium” and “high” attention tasks it would not appear in the first 3rd of letters to avoid users obtaining the color early enough to shift their attention to the periphery before the trial ended.

3.3 User Study

Setup. Due to the need to display high resolution stimuli across a wide field of view, we conduct our study using a 34 inch, 144 Hz Dell Curved Gaming Monitor (Model No. S3422DWG, see Fig. 2). This display has an adjustable backlight, allowing us to tune luminance. For this study we use a setting that gives a minimum and maximum

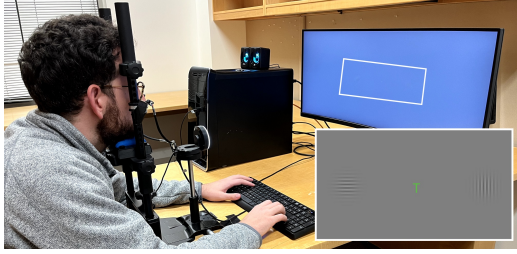


Fig. 2. Photograph of the user study setup. The inset shows an enlarged illustration of the stimulus on the screen; the central RSVP letter task with the Gabor patches centered at e to the left and right. The brightness of the letter T has been exaggerated for visibility.

luminance of 0.6 cd/m^2 and 104 cd/m^2 , respectively, and a gamma of 1.89. The neutral gray background triggered luminance adaptation to 28 cd/m^2 . A 2×2 spatial dithering was used to avoid visible color banding in the low-contrast stimuli. We used Python’s PsychoPy toolbox [Peirce 2007] and a custom shader to stream frames to the display by wired HDMI connection. All subjects were tested in a well-lit room and viewed the video display binocularly from an SR Research headrest situated 94 cm away, thus giving a field of view of $46^\circ \times 20^\circ$ and a resolution of 71 pixels per degree of visual angle. Pupil Labs Core eye trackers were mounted to the headrest to verify central gaze fixation throughout all studies.

Subjects. Ten adults participated (age range 23–29, 2 female). Due to the demanding nature of our psychophysical experiment, only a few subjects were recruited, which is common for similar low-level psychophysics (see e.g. [Patney et al. 2016]). All subjects in this and subsequent experiments had normal or corrected-to-normal vision, no history of visual deficiency, and no color blindness, but were not tested for peripheral-specific abnormalities. All subjects gave informed consent. The research protocol was approved by the Institutional Review Board at the host institution.

Procedure. To begin the study, subjects were set up in a comfortable position on the headrest and the eye tracker was calibrated using a 5-point screen calibration [Kassner et al. 2014]. The thresholds for each contrast condition (stimuli No. 1–3 in Table 1) were then estimated in a random order. For each condition, a two-alternative forced-choice (2AFC) adaptive staircase designed using QUEST [Watson and Pelli 1983] was used to measure the contrast discrimination threshold for each attention condition, starting with the “low”, then the “medium” and ending with the “high” foveal attention condition. At each step, the subject was shown a small (1°) white fixation cross for 1.2 s to indicate where they should fixate, followed by the attention-modulating task and contrast stimuli for 500 ms, then a Gaussian white noise screen for 1 s (to reduce after images). The subject was then given 10 s to indicate via different sets of marked buttons on a keyboard whether the target letter “T” was red or green, followed by whether the contrast patterns were of the same or different orientations. If the subject failed to answer during that time, the trial would be replayed. Each of the 3 test conditions at each of the 3 attention conditions were tested twice per subject,

taking approximately 90 minutes, with subjects encouraged to take breaks between staircases.

Results. Mean contrast thresholds across subjects are shown in Fig. 3a (see Supplement for table of values). It can be seen that the contrast thresholds are almost identical for the “low” attention condition, agreeing with the theory of *cortical magnification* described by Virsu and Rovamo [1979; 1979]. For the “medium” attention condition, however, the contrast thresholds do increase significantly with eccentricity ($p < 0.05$, paired t-test between neighboring eccentricities), almost $2 \times$ for 7° and over $3 \times$ for 21° (see Fig. 3b). Similarly, the “high” attention condition exhibits up to $4 \times$ threshold increase within our measured eccentricity range ($p < 0.05$). The increase in gain factors with eccentricity is consistent with work by Staugaard et al. [2016] who showed a decrease in attentional capacity with increasing stimulus eccentricity, when stimuli are scaled in size to account for cortical magnification. Furthermore, we observe significant differences between individual attention modes across the eccentricity ranges ($p < 0.01$ for most pairs, $p < 0.05$ for the “medium” and “high” attention, paired t-test with Bonferroni correction). This confirms our assumption that increasing the task difficulty will shift attention towards the fovea at a cost to sensitivity in the periphery. On the other hand, despite the considerable difference between the “low” and “medium” condition gradients, the gradients of the “medium” and “high” conditions are surprisingly similar, suggesting that the effect of attention modulation is non-linear.

3.4 Per-condition Model

The observed increase in thresholds for larger eccentricities is nearly linear with a small distortion which we describe using the square root of eccentricity to fit attention-dependent contrast threshold models:

$$t_a(e) = p_0 \sqrt{e} + p_1 \quad (1)$$

where a denotes one of our foveal attention conditions (“low”, “medium” or “high”). See Table 2 for parameters and Fig. 3a for plots.

Table 2. Fitted parameters of our attention-aware contrast threshold model $t_a(e)$ (Eq. 1). R^2 is the coefficient of determination.

a	p_0	p_1	R^2
“low”	$9.672 \cdot 10^{-4}$	$2.741 \cdot 10^{-2}$	0.705
“medium”	$2.737 \cdot 10^{-2}$	$-1.620 \cdot 10^{-2}$	1.000
“high”	$2.714 \cdot 10^{-2}$	$1.612 \cdot 10^{-2}$	0.956

As contrast sensitivity varies among observers, we are primarily interested in the relative attention gain represented by threshold elevations defined with respect to the “low” attention baseline condition as:

$$g_a(e) = \frac{t_a(e)}{t_{\text{low}}(e)} \quad (2)$$

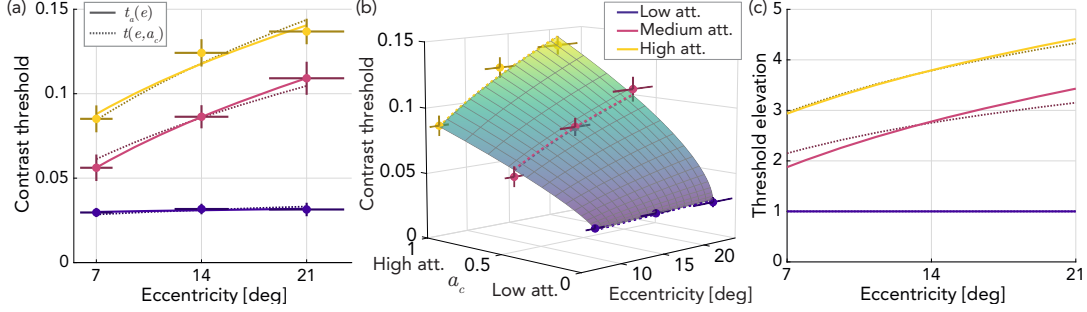


Fig. 3. Main study: (a) The mean measured contrast thresholds and the fitted attention curves for the per-condition model $t_a(e)$ (Eq. 1, full lines) and the unified model $t(e, a_c)$ (Eq. 4, dotted lines). The horizontal bars display extent of the Gabors. The vertical error bars show standard error. (b) A continuous attention-eccentricity fit of the unified model $t(e, a_c)$ (Eq. 4). (c) The attention gains $g_a(e)$ relative to the “low” foveal attention condition computed for each of the two models.

Assuming orthogonality of the attention effect and other independent parameters of the stimulus, we can formulate the attention-aware contrast sensitivity as:

$$S_a(e, \dots) = S(e, \dots) \frac{1}{g_a(e)} \quad (3)$$

where S is any of the CSF models discussed in Sec. 2.2. In Sec. 3.6 we use the StelaCSF [Mantiuk et al. 2022] model.

3.5 Unified model

Additionally, we explore a speculative model unifying the eccentricity e with a continuous interpretation of the attention condition $a_c \in [0, 1]$ where {“low” $\rightarrow 0$, “medium” $\rightarrow 0.5$, “high” $\rightarrow 1$ }. We design this model as an attention-dependent sweep between the per-attention curves, parameterized relative to our lowest eccentricity of 7° . We model the dependency for the slope and intercept separately using two gamma curves $a_c^{\gamma_s}$ and $a_c^{\gamma_i}$ to account for the non-linear perception of the different attention conditions. Due to the extreme non-linearity of the slope development we constrain γ_s to 0.5 and fit:

$$t(e, a_c) = \Psi \left(s_0, s_1, a_c^{\gamma_s} \right) \cdot \left(\sqrt{e} - \sqrt{7} \right) + \Psi \left(i_0, i_1, a_c^{\gamma_i} \right) \quad (4)$$

to our measured data. Here, $\{s_0, s_1, i_0, i_1, \gamma_i\} = \{0.00243, 0.0307, 0.0285, 0.0844, 0.771\}$ are the fitted parameters (DoF-adjusted $R^2 = 0.973$) and $\Psi(\alpha, \beta, w) = \alpha(1 - w) + \beta w$ is a linear interpolation function.

We compare the resulting unified model to our per-condition models $t_a(e)$ in Fig. 3a. Despite the lower parameter count, the unified model fits the measured data within the measurement errors. While the unified model allows for convenient interpolation, we argue for fitting task-specific models in practice, because the connection between the task and attention is highly individual and not well understood. Hence, we use our per-condition models $t_a(e)$ (Eq. 1) throughout the rest of this paper wherever not explicitly specified otherwise.

3.6 Validation

In Eq. 3, we apply attention correction as a multiplicative factor under an assumption of orthogonality between the two functions.

If this assumption holds, the difference between new thresholds predicted by our model and their measured values should be low.

We test this by measuring the attention gains for 4 new stimuli with different cortical magnifications and adaptation luminance levels than in our main study. We then compare the thresholds obtained by direct measurement with the thresholds predicted by our attention gain $g_a(e)$.

Experiment. We use the same experiment procedure as for the main study, except with 2 parameter sets used to fit StelaCSF [2022], a recently demonstrated unified model of CSF (see stimuli No. 4 and 5 in Table 1). These points were selected from the only 2 available datasets measured using stationary stimuli outside the fovea, with spatial frequency, eccentricity and size as different as possible to the stimuli used to fit our attention-aware CSF model. Additionally, we test effect of varying luminance adaptation on one of these datapoints by adjusting the backlight of our display (see stimuli No. 5–7).

Subjects. Eleven adults participated (age range 23–29, 5 female), six for stimuli No. 4 and 5 and five for stimuli No. 6 and 7. Only three of these subjects participated in the main study.

Results. We measured mean thresholds for the validation stimuli (No. 4–7) and the “low” attention condition as 0.032, 0.045, 0.57 and 0.51, which we use as baselines for a relative multiplicative adjustment of our measurements to corresponding predictions of StelaCSF. We compute Interquartile Range (IQR) of this multiplicative factor to detect outliers. We treat each of the 2 per-user repetitions as a single data sample and we remove a total of 3 strong outliers with offset of 4 or more IQR from the quartiles. We then apply this base adjustment consistently to all individual measurements to remove variability of the base sensitivity performance among users and instead focus on relative gains between attention conditions (see Fig 4b). The resulting adjusted measurements are then compared with the contrasts predicted by the original attention-unaware StelaCSF model and our derived attention-aware CSF model $S_a(e, \dots)$. We compute the error of both models in Fig. 4c.

For the 2 stimuli isoluminant with our model data (No. 4 and 5), we observe statistically significantly lower error between our and the original StelaCSF predictions with respect to the experimentally

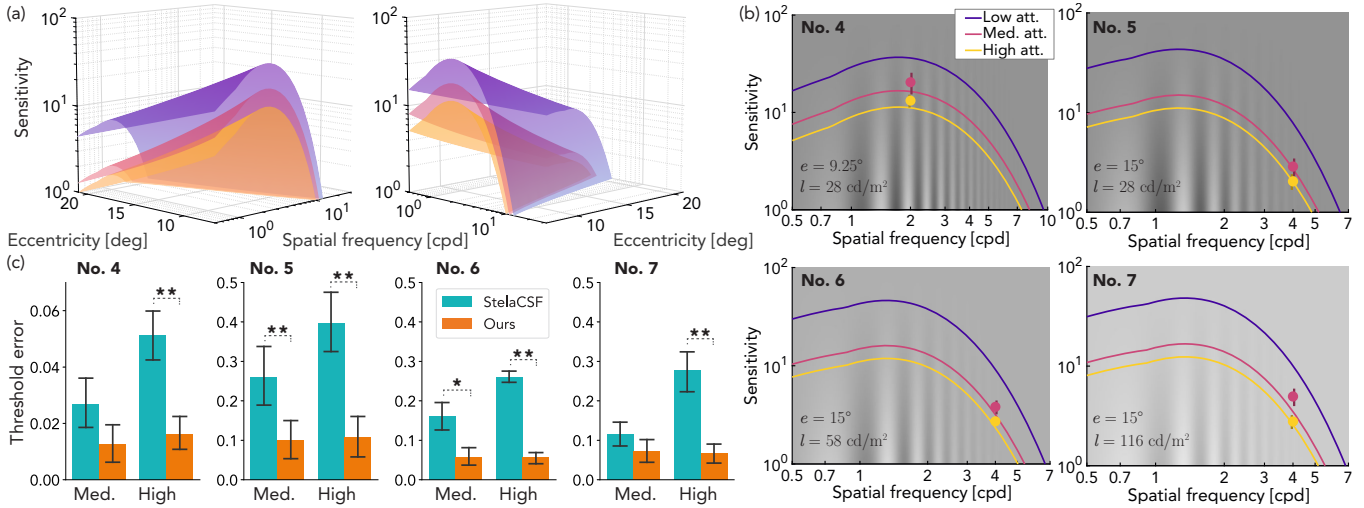


Fig. 4. Validation studies: (a) Two different views of an eccentricity vs. spatial frequency plot for the original StelaCSF [Mantiuk et al. 2022] model (the top surface, in purple) and our scaled models $S_{\text{medium}}(e, f_s)$ (in magenta), $S_{\text{high}}(e, f_s)$ (in yellow) for a static stimulus with an area of 1 deg^2 and an adaptation luminance of 28 cd/m^2 (same as our model study in Sec. 3.3). (b) Slices of the same models describing dependency on spatial frequency for the conditions used in our validation study (see No. 4–7 in Table 1). The points denote directly measured sensitivities scaled relative to the baseline. The bars are 95% confidence intervals. (c) Corresponding threshold prediction errors of StelaCSF vs. our model (lower is better). The error bars are 95% confidence intervals and significance is indicated at the $p < 0.05$ and 0.01 levels with * and ** respectively (Wilcoxon test).

measured thresholds in all conditions except for one. The lower observed difference between “low” and “medium” attention for the lower frequency stimulus (No. 4) points to an overestimation of the gain by our model here. Notably, even in this worst case, the prediction error is still lower than that of the baseline StelaCSF.

As a practical example, our measurements indicate that with “high” attention and a 28 cd/m^2 display, a spatial pattern with $f_s = 2 \text{ cpd}$ shown at eccentricity of 15° will be just discriminable if rendered with an amplitude value of 32 (for a 0–255 signal range of an 8-bit display with gamma of 2.2) while our model would yield amplitude of 30 and the baseline model would adhere to the original stelaCSF prediction of amplitude 8.

The favorable performance of our model also holds for the other 2 luminance levels (stimuli No. 6 and 7). Despite this, we observe a trend of attention gain reduction with increasing luminance which is significant for the “medium” attention at 116 cd/m^2 ($g_a : 3.03 \rightarrow 2.15$, $p < 0.05$, Mann-Whitney U test) and “high” attention at 58 cd/m^2 ($g_a : 4.12 \rightarrow 3.37$). This compression could be caused by the overall increase of sensitivity under such conditions and should be considered by users of our model.

To summarize, our experiment suggests that while the assumption of full orthogonality is unlikely to hold everywhere, the relative benefit of including the attention model may still be stronger relative to the cost of this simplification.

4 ATTENTION-AWARE FOVEATED RENDERING

The goal of foveated rendering is to reduce computational cost without introducing perceptible artifacts by exploiting the reduction of vision performance in the periphery, typically by adjusting sampling rate with respect to peripheral acuity drop (see Sec. 2.1). The quality

of such foveation can be assessed by visual difference predictors, for example, FovVideoVDP [Mantiuk et al. 2021], a state-of-the-art metric that models the spatial, temporal, and peripheral aspects of perception. In this section, we experimentally validate whether integration of our attention-aware perceptual model improves the performance of FovVideoVDP in predicting visibility of foveation artifacts under varying attention conditions. To that end, we emulate a simple foveated renderer and separately calibrate the foveation intensity for three different attention regimes in a user study. We then compare the perceptual errors of the calibrated stimuli predicted by FovVideoVDP with and without our attention-aware model to assess their agreement with human judgment.

4.1 Measuring imperceptible foveation

Similar to the contrast discrimination task in Sec. 3, we create a space-multiplexed comparison of foveated images. In particular, we split the screen into left and right sides, apply foveation to one side only (randomly selected) and ask subjects which of the sides appeared more visually degraded (see Fig. 5a). Then, by modifying the parameters of the foveated side, we can find the threshold for which the foveation is nearly imperceptible to the subject. The central transition around the fixation was replaced by a neutral background vertical bar with 6° width and Gaussian fall-off (standard deviation of 0.5°) and the attention-modulating RSVP task, as in Sec. 3.2, was then displayed centrally.

For the foveation, we base our approach on the work of Guenter et al. [2012] and the linear *minimum angle of resolution* (MAR) model describing the reciprocal of acuity as:

$$\omega(e) = me + \omega_0 \quad (5)$$

where the bias $\omega_0 = 1/48^\circ$, as in the original work, and the slope m is a free variable measured as a threshold in our study. The peripheral resolution decrease was simulated by an approximated Gaussian filter with spatially varying standard deviation:

$$\sigma(e) = \frac{\omega/\omega_s - 1}{2\sigma_c} \quad (6)$$

where $\omega_s = 0.0283^\circ$ is the peak MAR of our display and $\sigma_c = 2$ is the chosen cut-off determining the assumed bandwidth of the filter. Note that this particular choice primarily affects the absolute value of our slopes and not the relative ratios between conditions.

4.2 User study

Setup. We used the same experimental setup as in Sec. 3.1. The stimuli consisted of one of four foveated images displayed across the entire screen with the gaze fixation directed to the center.

Subjects. Thirty adults participated (age range 23–29, 10 female). Subjects were first shown the original images, to avoid exploratory saccades during the study, and then shown an example of the foveation effect.

Procedure. As in the studies in Sec. 3.1, subjects were instructed to always fixate on the central RSVP task and observe the foveation task concurrently in their periphery. The thresholds for each image were estimated in a random order for each subject. For each image, a 2-AFC adaptive staircase using QUEST [Watson and Pelli 1983] was used to measure the threshold of the foveation slope m for each attention condition, starting with the “low”, followed by the “medium” and ending with the “high” foveal attention condition. Similar to the previous studies, at each step, the subject is shown the attention-modulating task and foveation detection task for 500 ms. The subject then indicated the color of the target letter “T”, and if correct, were asked which side of the image (left or right) was more visually degraded. Whenever the subject incorrectly answered the RSVP task, they were forced to start the step again. Each subject viewed 2 images, either “Tulips” and “City” or “Mountain” and “Forest” (see Fig. 5a) at each of the 3 attention conditions, to keep the study duration to approximately 45 minutes (including breaks).

Results. In Fig. 5d, we show the measured MAR slopes m averaged across the users (labeled “Measured”). We applied the same IQR procedure as in Sec. 3.6 but we did not detect any outliers. As expected, for all images the slope significantly increases ($p < 0.001$, paired t-test with Bonferroni correction) for both “medium” and “high” compared to “low” attention conditions. This means that a more aggressive foveation becomes acceptable as attention shifts from periphery towards fovea. Furthermore, we note that there are statistically significant differences between slopes measured for at least some image pairs with “low” (one-way ANOVA, $F(3, 116) = 4.99$, $p = 0.003$), “medium” ($F(3, 116) = 10.26$, $p < 0.001$) and “high” ($F(3, 116) = 3.87$, $p = 0.011$) attention. This points to a content-dependent nature of the problem. In the next section, we discuss whether a visual difference predictor could be used to predict foveation parameters for a specific image.

4.3 Foveated quality prediction

Setup. Acuity-driven foveation algorithms conservatively account for the worst-case scenario of the smallest detectable image detail [Tariq et al. 2022]. As seen in our results, distribution of contrast in specific images affects the acceptable foveation intensity. This is modeled by visual difference predictors such as FovVideoVDP [Mantiuk et al. 2021], which decomposes an image into spatio-temporal frequency bands and models their visibility by utilizing the CSF and a contrast masking model. However, while explicitly modeling retinal eccentricity, the original CSF does not account for attention. We experimentally modify the authors’ implementation and integrate our model as an orthogonal scaling factor of the CSF component. We then apply the original and the modified predictor to assess the quality of the foveated images with the per-image calibrated slopes from Sec. 4.2. Furthermore, we evaluate whether an inverse process could be used to optimize the foveation intensity.

Quality metric. In Fig. 5b, we display quality scores produced by the original FovVideoVDP metric compared to our modified predictor obtained by computing visual difference between a foveated image calibrated by each individual subject and the full-quality reference. The scores were averaged for each image, for each of the “medium” and “high” attention conditions. We compare these to the expected value (labeled “Measured”) which was obtained by FovVideoVDP for foveated images calibrated with the “low” attention condition. We assume that this represents the personal subject-specific threshold of the perceived quality for the given image and that it should remain constant under varying attention.

Following our previous results, we expect that images with larger objective degradation should be judged as having equivalent quality and that a successful prediction should reflect that. Consequently, we observe that the error measured as a relative difference between our prediction and the “Measured” value is consistently lower than that for FovVideoVDP ($p < 0.001$, Wilcoxon test). This suggests that our modified predictor is better aligned with the attention-modulated perception.

In Fig. 5c, we additionally compare maps of Just-Objectionable-Differences (JOD) produced by both predictors for the mean calibrated slopes at each condition. FovVideoVDP indicates a strong increase of perceived artifacts even as attention towards the fovea (“high” attention condition). Our method instead predicts errors on the boundary of visibility for all conditions which is consistent with our assumption.

Finally, we note that the “low” attention score predicted by FovVideoVDP based on the directly measured slopes is significantly different between at least some of the images (9.375 for “Tulips”, 8.383 for “City”, 9.470 for “Mountain” and 9.314 for “Forest”, $F(3, 116) = 150.5$, $p < 0.001$, one-way ANOVA). Since this is the baseline condition, this discrepancy is orthogonal to our primary objective of exploring the overall impact of attention, and thus we defer its investigation to future work.

MAR slope prediction. The visual difference prediction potentially allows us to optimize foveation parameters by posing it as a constrained problem:

$$\Theta = \arg \min_{\Theta} C(\Theta) \quad \text{subject to} \quad Q(\Theta) \geq Q_{\text{thr}} \quad (7)$$

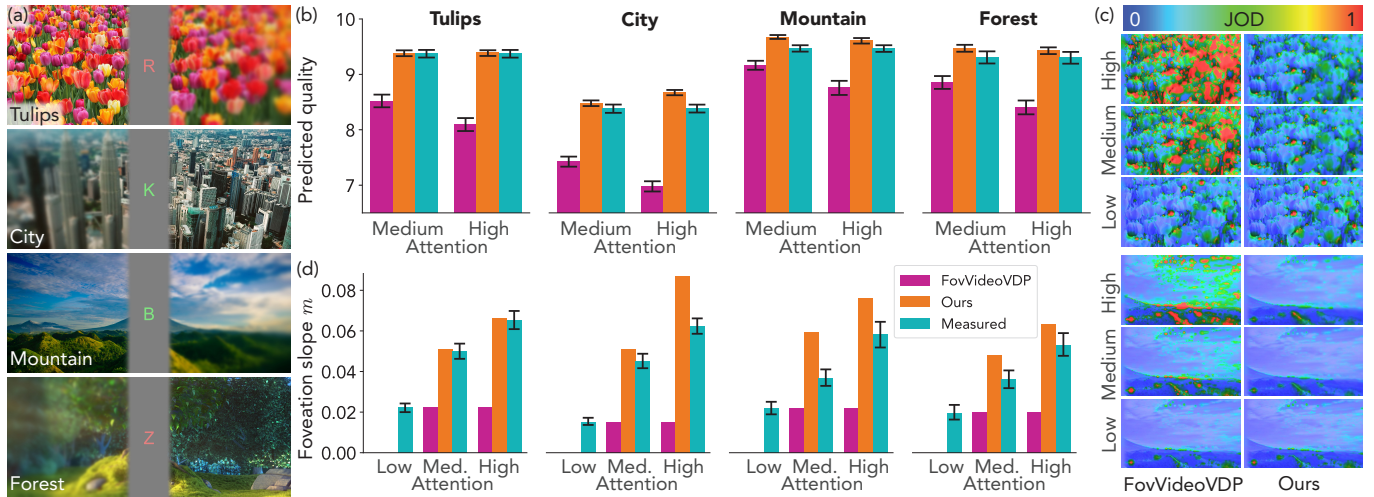


Fig. 5. Foveation study: (a) Stimuli from our study showing the attention-modulating RSVP task in fovea and the peripheral foveation detection task. One side (randomly selected) is foveated while the other is left at full resolution. The foveation effect and the color and size of the RSVP task are exaggerated for visibility. (b) Quality scores predicted by the original FovVideoVDP metric vs. our modified predictor (closer to *Measured* is better) for the foveated images calibrated in our user study. The error bars show 95% confidence intervals. The *Measured* quality refers to the actual quality threshold measured in the “low” condition. (c) Comparison of visual difference maps produced by the original FovVideoVDP metric vs. our modified predictor for the calibrated MAR slopes. Colors visualize Just-Objectable-Differences (JOD) with respect to the original “Tulips” and “Mountain” images (small section from the right periphery shown). (d) Comparison of MAR slopes (intensities) predicted by the original FovVideoVDP metric vs. our modified predictor for each image compared to the *Measured* slopes (closer to *Measured* is better). The legend is shared with panel (b). The error bars show 95% confidence intervals of the measured values. Note that the model-based slope predictors do not yield variance (no error bars shown).

where Θ is a set of rendering parameters, $C(\Theta)$ is the cost of the rendering (typically time and power consumption), $Q(\Theta)$ is an image quality predictor and Q_{thr} the required threshold. In our case $\Theta = m$, $C(\Theta)$ is a monotonically decreasing function of m , $Q(\Theta)$ is provided by our visual difference predictor (with access to the reference image) and Q_{thr} is obtained from the “low” foveal attention condition in Sec. 4.2. The resulting problem of one variable can be efficiently solved by bisection.

Ideally, we could use a single Q_{thr} for any image. However, due to the significant difference between “low” attention thresholds obtained for our images, we opt to use scene specific Q_{thr} of 9.375 for “Tulips”, 8.383 for “City”, 9.470 for “Mountain” and 9.314 for “Forest”. This simulates a correction function that calibrates the underlying predictor for content-dependent effects and development of which is outside of the scope of this work.

Fig. 5d compares the MAR slopes obtained by solving the inverse problem with $Q(\Theta)$ implemented using the original FovVideoVDP metric and our modified predictor. While our predicted slopes do not always match the directly measured values, for the “high” attention the errors are consistently lower than those from FovVideoVDP ($p < 0.05$ for the “Mountain”, $p < 0.01$ for the rest, non-parametric Wilcoxon test). This is remarkable given the large domain gap between the model and foveation stimuli. It suggests that our model is useful for attention-aware foveated rendering. Similarly, we observe statistically lower prediction errors of our model with the “medium” attention for the “Tulips” and “City” images ($p < 0.01$) while no statistically significant differences were measured for the rest.

The remaining error could originate from a multitude of sources, among them the orthogonal assumption of CSF scaling as well as other higher level effects not accounted for in either model. To illustrate the impact, in the worst case of the “City” image with “high” attention and the extreme periphery of 46° in our experimental display setup, this error would lead to removal of spatial details in the 0.25–0.35 cpd band which might be noticeable based on our measured data.

Importantly, we observe that the bias towards overestimation of the slopes is consistent. This hints to feasibility of fine tuning for a specific foveation algorithm. Even without such treatment, the relative preference of our model over the baseline is most prominent for the “high” attention which is particularly relevant for many applications where users focus at a specific target on the screen.

4.4 Bandwidth analysis

In this section we analyze the additional computation gain that can theoretically be obtained by using our attention-aware model when the user is focused on a task in the fovea. While we could analyze the bandwidth by decomposing the image into frequency bands and discarding the signal following the CSF predictions, we decided on a more conservative approach that instead uses our direct perceptual measurements of vision performance under the specific foveal (RSVP) task. Therefore, we model the foveation algorithm by Guenter et al. [2012] together with the global mean MAR slopes m obtained for the “low”, “medium” and “high” attention conditions as 0.0198, 0.0420 and 0.0596.

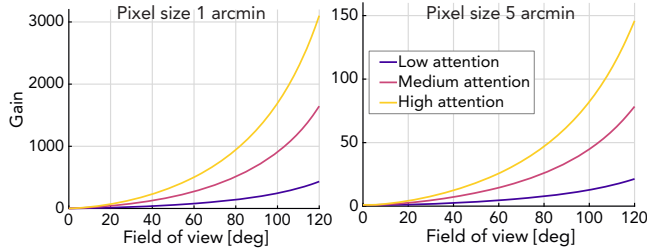


Fig. 6. Computational gain analysis as a fraction of original and retained pixel sampling density depending on pixel size and the covered visual field (horizontal axis). Note the difference in the gain axes scales.

Unlike the discrete segmentation in the original algorithm, we simplify the analysis by assuming that sampling rate of each pixel can be controlled independently and hence we directly map the local MAR $\omega(e)$ (Eq. 5) to the computational gain Ψ derived from local areal sampling density as:

$$\Psi(\text{FOV}) = \left(\int_{\text{FOV}} 1 \, dx \right) \cdot \left(\int_{\text{FOV}} \max \left(\frac{\omega(x)}{\omega_s}, 1 \right)^{-2} dx \right)^{-1} \quad (8)$$

where ω_s is the peak MAR of a given display with a two dimensional field of view FOV.

In Fig. 6, we display computational gains obtained as a function of display field of view for a common 20 ppd (pixels per degree) and future high-density 60 ppd displays as an upper bound for the analyzed algorithm. It must be noted that gains in real applications are influenced by efficiency of a particular renderer. As our contributions are independent of such design choices, more advanced foveation approaches such as noise-based enhancement [Tariq et al. 2022] can be considered for additional gains.

5 DISCUSSION

The experimental data we measure and the models we fit to them further our understanding of human perception and lay the foundation of future attention-aware foveated graphics techniques. Yet, several important questions remain to be discussed.

Limitations and Future Work. While our studies clearly demonstrate that modulating attention distribution between the periphery and fovea strongly impacts contrast sensitivity and foveation efficacy, we do not propose a method to measure attention. In contrast to overt attention, which is readily measurable using eye tracking, covert attention is much more challenging. A promising direction for exploration is the relation between pupil dilation and attentional effort [Hoeks and Levelt 1993] and in some scenarios pupillary light response [Mathôt et al. 2013]. One might also investigate the combination of eye tracking with image salience [Itti and Koch 2000] or other metrics. It should also be noted that training and practice can significantly improve the ability to split attention between the fovea and periphery [Zhang et al. 2022]. This could lead to decrease in attention dedicated to the foveal RSVP task. To mitigate this, we randomize trial order and encourage sufficient rest time, yet such an approach is time consuming and limits the CSF gamut that can be

measured in one sitting. While we show that our orthogonal scaling approach still leads to favorable performance when compared to baselines, we emphasize that our attention model should not be extrapolated outside of the measured eccentricities. Finally, rather than proposing a novel foveation algorithm, we focus on demonstration of the perceptual effect as a whole. Future work should investigate more advanced foveation algorithms and explore the effect of attention in the temporal domain.

Conclusion. At the convergence of applied vision science, computer graphics, and wearable computing system design, foveated graphics techniques will play an increasingly important role in future VR/AR systems. With our work, we hope to motivate the importance of cognitive science in human perception and inspire a new axis of approaches within foveated graphics.

ACKNOWLEDGMENTS

The project was supported by a Stanford Knight-Hennessy Fellowship and Samsung. The authors would also like to thank Robert Konrad, Nitish Padmanaban and Justin Gardner for helpful discussions and advice.

REFERENCES

- Katharina Anton-Erxleben and Marisa Carrasco. 2013. Attentional enhancement of spatial resolution: linking behavioural and neurophysiological evidence. *Nature Reviews Neuroscience* 14, 3 (2013), 188–200.
- E Leslie Cameron, Joanna C Tai, and Marisa Carrasco. 2002. Covert attention affects the psychometric function of contrast sensitivity. *Vision research* 42, 8 (2002), 949–967.
- Marisa Carrasco. 2011. Visual attention: The past 25 years. *Vision research* 51, 13 (2011), 1484–1525.
- Marisa Carrasco. 2018. How visual spatial attention alters perception. *Cognitive processing* 19, 1 (2018), 77–88.
- Marisa Carrasco, Anna Marie Giordano, and Brian McElree. 2006. Attention speeds processing across eccentricity: Feature and conjunction searches. *Vision research* 46, 13 (2006), 2028–2040.
- Marisa Carrasco, Cigdem Penpeci-Talgar, and Miguel Eckstein. 2000. Spatial covert attention increases contrast sensitivity across the CSF: support for signal enhancement. *Vision research* 40, 10-12 (2000), 1203–1215.
- Nianchen Deng, Zhenyi He, Jiannan Ye, Budmonde Duinkharjav, Praneeth Chakravarthula, Xubo Yang, and Qi Sun. 2022. FoV-NeRF: Foveated Neural Radiance Fields for Virtual Reality. *IEEE Transactions on Visualization and Computer Graphics* 28, 11 (2022), 3854–3864.
- Andrew T. Duchowski, Nathan Cournia, and Hunter A. Murphy. 2004. Gaze-Contingent Displays: A Review. *Cyberpsychology & behavior* 7 (2004), 621–34.
- Budmonde Duinkharjav, Kenneth Chen, Abhishek Tyagi, Jiayi He, Yuhao Zhu, and Qi Sun. 2022. Color-Perception-Guided Display Power Reduction for Virtual Reality. *ACM Transactions on Graphics (TOG)* 41, 6 (2022), 1–16.
- Charles W Eriksen and James D St. James. 1986. Visual attention within and around the field of focal attention: A zoom lens model. *Perception & psychophysics* 40, 4 (1986), 225–240.
- Sebastian Friston, Tobias Ritschel, and Anthony Steed. 2019. Perceptual Rasterization for Head-mounted Display Image Synthesis. *ACM Trans. Graph. (Proc. SIGGRAPH 2019)* 38, 4 (2019).
- Wilson S. Geisler and Jeffrey S. Perry. 1998. Real-time foveated multiresolution system for low-bandwidth video communication. In *Human Vision and Electronic Imaging III*, Vol. 3299. International Society for Optics and Photonics, 294–305.
- Brian Guenter, Mark Finch, Steven Drucker, Desney Tan, and John Snyder. 2012. Foveated 3D Graphics. *ACM Trans. Graph. (SIGGRAPH Asia)* (2012).
- Bert Hoeks and Willem JM Levelt. 1993. Pupillary dilation as a measure of attention: A quantitative system analysis. *Behavior Research methods, instruments, & computers* 25, 1 (1993), 16–26.
- Liqliang Huang and Karen R Dobkins. 2005. Attentional effects on contrast discrimination in humans: evidence for both contrast gain and response gain. *Vision research* 45, 9 (2005), 1201–1212.
- Laurent Itti and Christof Koch. 2000. A saliency-based search mechanism for overt and covert shifts of visual attention. *Vision research* 40, 10-12 (2000), 1489–1506.
- Anton S. Kaplanyan, Anton Sochenov, Thomas Leimkühler, Mikhail Okunev, Todd Goodall, and Gizem Rufo. 2019. DeepFovea: Neural Reconstruction for Foveated

- Rendering and Video Compression Using Learned Statistics of Natural Videos. *ACM Trans. Graph.* 38, 6, Article 212 (Nov. 2019), 13 pages.
- Moritz Kassner, William Patera, and Andreas Bulling. 2014. Pupil: an open source platform for pervasive eye tracking and mobile gaze-based interaction. In *Proceedings of the 2014 ACM international joint conference on pervasive and ubiquitous computing: Adjunct publication*. 1151–1160.
- D. H. Kelly. 1979. Motion and vision. II. Stabilized spatio-temporal threshold surface. *JOSA* 69, 10 (Oct. 1979), 1340–1349.
- George Alex Koulteris, Kaan Akşit, Michael Stengel, Rafał K Mantiuk, Katerina Mania, and Christian Richardt. 2019. Near-eye display and tracking technologies for virtual and augmented reality. In *Computer Graphics Forum*, Vol. 38. Wiley Online Library, 493–519.
- Brooke Krajancich, Petr Kellnhofer, and Gordon Wetzstein. 2021. A perceptual model for eccentricity-dependent spatio-temporal flicker fusion and its applications to foveated graphics. *ACM Transactions on Graphics (TOG)* 40, 4 (2021), 1–11.
- Dale K Lee, Christof Koch, and Jochen Braun. 1997. Spatial vision thresholds in the near absence of attention. *Vision research* 37, 17 (1997), 2409–2418.
- Sam Ling and Marisa Carrasco. 2006. Sustained and transient covert attention enhance the signal via different contrast response functions. *Vision research* 46, 8-9 (2006), 1210–1220.
- David Luebke and Benjamin Hallen. 2001. Perceptually driven simplification for interactive rendering. In *Proceedings of the 12th Eurographics conference on Rendering (EGWR'01)*. Eurographics Association, Goslar, DEU, 223–234.
- Arien Mack. 2003. Inattentive blindness: Looking without seeing. *Current directions in psychological science* 12, 5 (2003), 180–184.
- Monireh Mahjoob and Andrew J Anderson. 2019. Contrast discrimination under task-induced mental load. *Vision Research* 165 (2019), 84–89.
- Monireh Mahjoob, Javad Heravian Shandiz, and Andrew J Anderson. 2022. The effect of mental load on psychophysical and visual evoked potential visual acuity. *Ophthalmic and physiological optics* 42, 3 (2022), 586–593.
- Rafał K. Mantiuk, Maliha Ashraf, and Alexandre Chapiro. 2022. StelaCSF: A Unified Model of Contrast Sensitivity as the Function of Spatio-Temporal Frequency, Eccentricity, Luminance and Area. *ACM Trans. Graph.* 41, 4, Article 145 (jul 2022), 16 pages. <https://doi.org/10.1145/3528223.3530115>
- Rafał K Mantiuk, Gyorgy Denes, Alexandre Chapiro, Anton Kaplanyan, Gizem Rufo, Romain Bachy, Trisha Lian, and Anjul Patney. 2021. Fovvideovdp: A visible difference predictor for wide field-of-view video. *ACM Transactions on Graphics (TOG)* 40, 4 (2021), 1–19.
- Rafał K. Mantiuk, Minjung Kim, Maliha Ashraf, Qiang Xu, Ming Ronnier Luo, Jasna Martinovic, and Sophie M. Wuerger. 2020. Practical Color Contrast Sensitivity Functions for Luminance Levels up to 10000 cd/m2.
- Sebastiaan Mathôt, Lotje Van der Linden, Jonathan Grainger, and Françoise Vitu. 2013. The pupillary light response reveals the focus of covert visual attention. *PLoS one* 8, 10 (2013), e78168.
- John D. McCarthy, M. Angela Sasse, and Dimitrios Miras. 2004. Sharp or smooth? comparing the effects of quantization vs. frame rate for streamed video. In *Proceedings of the SIGCHI Conference on Human Factors in Computing Systems (CHI '04)*. Association for Computing Machinery, New York, NY, USA, 535–542.
- Barbara Montagna, Franco Pestilli, and Marisa Carrasco. 2009. Attention trades off spatial acuity. *Vision research* 49, 7 (2009), 735–745.
- Maria Concetta Morrone, V Denti, and D Spinelli. 2004. Different attentional resources modulate the gain mechanisms for color and luminance contrast. *Vision research* 44, 12 (2004), 1389–1401.
- Hunter Murphy and Andrew T. Duchowski. 2001. Gaze-Contingent Level Of Detail Rendering. (2001).
- 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 Transactions on Graphics (TOG)* 35, 6 (Nov. 2016), 179:1–179:12.
- Jonathan W Peirce. 2007. PsychoPy—psychophysics software in Python. *Journal of neuroscience methods* 162, 1-2 (2007), 8–13.
- John G Robson. 1966. Spatial and temporal contrast-sensitivity functions of the visual system. *Josa* 56, 8 (1966), 1141–1142.
- Jyrki Rovamo and Veijo Virsu. 1979. An estimation and application of the human cortical magnification factor. *Experimental brain research* 37, 3 (1979), 495–510.
- George Sperling and Melvin J Melchner. 1978. The attention operating characteristic: Examples from visual search. *Science* 202, 4365 (1978), 315–318.
- Camilla Funch Staugaard, Anders Petersen, and Signe Vangkilde. 2016. Eccentricity effects in vision and attention. *Neuropsychologia* 92 (2016), 69–78.
- Qi Sun, Fu-Chung Huang, Joohwan Kim, Li-Yi Wei, David Luebke, and Arie Kaufman. 2017. Perceptually-guided foveation for light field displays. *ACM Transactions on Graphics* 36, 6 (Nov. 2017), 192:1–192:13.
- Taimoor Tariq, Cara Tursun, and Piotr Didyk. 2022. Noise-based enhancement for foveated rendering. *ACM Transactions on Graphics (TOG)* 41, 4 (2022), 1–14.
- Cara Tursun and Piotr Didyk. 2022. Perceptual Visibility Model for Temporal Contrast Changes in Periphery. *ACM Trans. Graph.* 42, 2, Article 20 (nov 2022), 16 pages. <https://doi.org/10.1145/3564241>
- Okan Tarhan Tursun, Elena Arabadzhyska-Koleva, Marek Wernikowski, Radosław Mantiuk, Hans-Peter Seidel, Karol Myszkowski, and Piotr Didyk. 2019. Luminance-contrast-aware foveated rendering. *ACM Transactions on Graphics (TOG)* 38, 4 (2019), 1–14.
- V Virsu and J Rovamo. 1979. Visual resolution, contrast sensitivity, and the cortical magnification factor. *Experimental brain research* 37, 3 (1979), 475–494.
- Andrew B Watson. 2018. The Field of View, the Field of Resolution, and the Field of Contrast Sensitivity. *Journal of Perceptual Imaging* 1, 1 (2018), 10505–1.
- Andrew B Watson and Albert J Ahumada. 2016. The pyramid of visibility. *Electronic Imaging* 2016, 16 (2016), 1–6.
- Andrew B Watson and Denis G Pelli. 1983. QUEST: A Bayesian adaptive psychometric method. *Perception & psychophysics* 33, 2 (1983), 113–120.
- MJ Wright and A Johnston. 1983. Spatiotemporal contrast sensitivity and visual field locus. *Vision research* 23, 10 (1983), 983–989.
- Qi Zhang, Zhibang Huang, Liang Li, and Sheng Li. 2022. Visual search training benefits from the integrative effect of enhanced covert attention and optimized overt eye movements. *bioRxiv* (2022).

SUPPLEMENTARY MATERIAL

In this section we provide additional discussion and results in support of the primary text.

Gabor patches in display space

In Section 3.1 of the main paper we discuss the use of Gabor patches for measuring CSF. A Gabor patch is a complex sinusoid modulated by a Gaussian envelope, defined as:

$$g(\mathbf{x}, \mathbf{x}_0, \theta, \sigma, f_s) = \exp\left(\frac{-\|\mathbf{x} - \mathbf{x}_0\|^2}{2\sigma^2}\right) \cos(2\pi f_s \mathbf{x} \cdot [\cos \theta, \sin \theta]), \quad (9)$$

where \mathbf{x} denotes the spatial location on the display, \mathbf{x}_0 is the center of the patch, σ is the standard deviation of the Gaussian in visual degrees, and f_s and θ are the spatial frequency in cpd and angular orientation in degrees for the sinusoidal grating function.

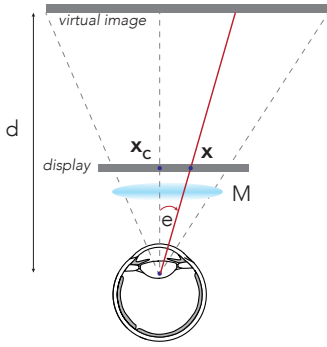


Fig. 7. Schematic illustrating the geometrical relationship used to convert spatial location in pixels to eccentricity in degrees of visual angle.

In our application, it is more convenient to describe spatial position in terms of eccentricity, measured in degrees of visual angle. Such transformation requires information about physical size of a pixel, p , and the dimensions and pixel resolution of the display (see Fig. 7). Then we can derive that:

$$\tan(e) = \frac{p(\mathbf{x} - \mathbf{x}_C)}{d/M}, \quad (10)$$

where e is eccentricity, \mathbf{x}_C is the location of the pixel directly in front of the eye, d is the distance to the virtual image and M is the magnification factor of the lenses. Finally, Eq. 10 can be re-arranged and substituted into Eq. 9 to re-define the Gabor function in terms of eccentricity.

The cortical magnification factor

It has been argued that the changes in detection and discrimination thresholds across eccentricity can be explained by the concept of *cortical magnification* [Rovamo and Virsu 1979; Virsu and Rovamo 1979]. This model describes how many neurons in the visual cortex are responsible for processing a particular part of the visual field. The central, foveal, region is processed by many more neurons (per

degree of visual angle) than the periphery. The cortical magnification is expressed in millimeters of cortical surface per degree of visual angle. We rely on the model by Dougherty et al. [?] (also used by FovVideoVDP [2021]) which was fitted to fMRI measurements of V1. It is modeled as:

$$M(e) = \frac{a_0}{e + e_2}$$

where e is eccentricity in visual degrees and the fitted parameters are $a_0 = 29.2$ mm and $e_2 = 3.67^\circ$.

Virsu and Rovamo [1979; 1979] showed that the differences in detection of sinusoidal patterns and also discrimination of their orientation or direction of movement, can be compensated by increasing the size of the stimuli in the peripheral vision and the size increase is consistent with the inverse of cortical magnification. In Section 3.1 of the main paper, we follow that observation to scale our stimuli to other retinal positions such for 2 cpd and diameter of 5° at 21° eccentricity, the corresponding spatial frequency (f_s) and area (A) can be calculated for other eccentricity (e) as $f_s = 2 \cdot \frac{M(e)}{M(21)}$ and $A = 5 \cdot \frac{M(21)}{M(e)}$.

Study results

We provide more details for the results of our experiments. Fig. 8 shows trends from the main study plotted for individual users (averaged across the two repetitions). Tables 3 and 4 list our measured thresholds for each stimulus and measured slopes for each image.

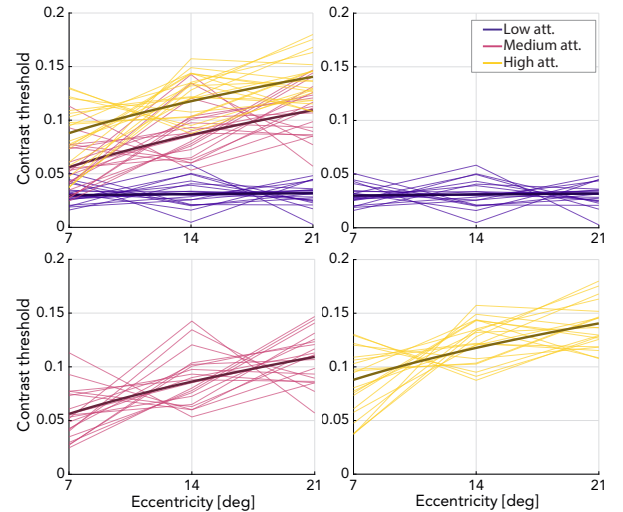


Fig. 8. Contrast thresholds measured for individual subjects (thin lines) in our main study that were used to fit our model (thick lines). For clarity, the attention levels are plotted together in the first panel and separately in the other panels. Mean thresholds for each plot line were rescaled to match the respective global attention level means in order to remove subject-specific variation of the base sensitivity and highlight the variation among attention levels and eccentricities.

Table 3. Mean measured contrast thresholds for the stimuli in our main and validation studies for the different attention conditions (Sec. 3.3 and Sec. 3.5 of the main paper).

No.	Contrast threshold		
	Low	Medium	High
1	0.0297	0.0561	0.0851
2	0.0317	0.0864	0.1242
3	0.0314	0.1091	0.1368
4	0.0325	0.0607	0.0905
5	0.0452	0.1304	0.1806
6	0.0573	0.1415	0.1926
7	0.0508	0.1059	0.1832

Table 4. Mean measured MAR slopes from the foveation study for the different attention conditions (Sec. 4.2 of the main paper).

Image	MAR slope		
	Low	Medium	High
Tulips	0.0222	0.0499	0.0651
City	0.0153	0.0449	0.0623
Mountain	0.0221	0.0369	0.0581
Forest	0.0197	0.0361	0.0531