stelaCSF - A Unified Model of Contrast Sensitivity as the Function of Spatio-Temporal Frequency, Eccentricity, Luminance and Area

RAFAŁ K. MANTIUK, University of Cambridge, UK
MALIHA ASHRAF, University of Liverpool, UK
ALEXANDRE CHAPIRO, Reality Labs, USA

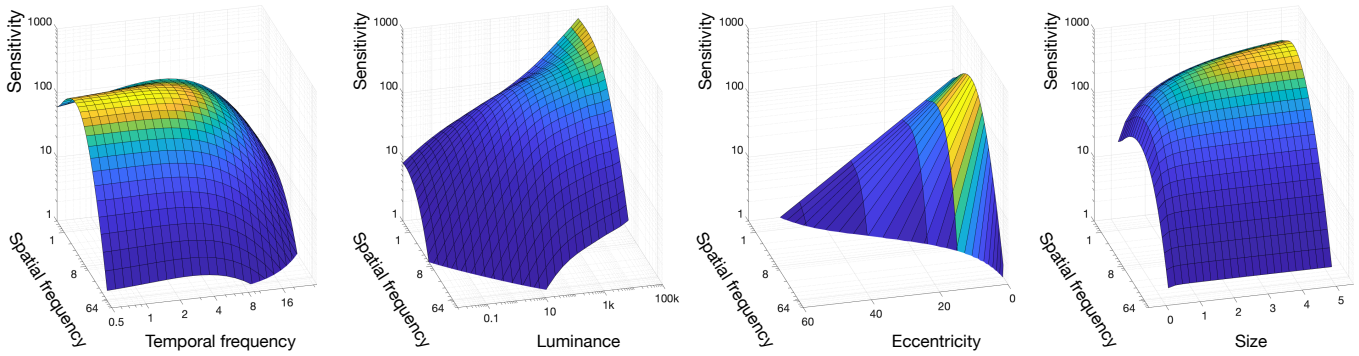

Fig. 1. Here, we visualize 3D slices of our novel 5-dimensional contrast sensitivity function, showing interactions between spatial frequency and four other factors (from left to right): temporal frequency, luminance, eccentricity and size.

A contrast sensitivity function, or CSF, is a cornerstone of many visual models. It explains whether a contrast pattern is visible to the human eye. The existing CSFs typically account for a subset of relevant dimensions describing a stimulus, limiting the use of such functions to either static or foveal content but not both. In this paper, we propose a unified CSF, stelaCSF, which accounts for all major dimensions of the stimulus: spatial and temporal frequency, eccentricity, luminance, and area. To model the 5-dimensional space of contrast sensitivity, we combined data from 11 papers, each of which studied a subset of this space. While previously proposed CSFs were fitted to a single dataset, stelaCSF can predict the data from all these studies using the same set of parameters. The predictions are accurate in the entire domain, including low frequencies. In addition, stelaCSF relies on psychophysical models and experimental evidence to explain the major interactions between the 5 dimensions of the CSF. We demonstrate the utility of our new CSF in a flicker detection metric and in foveated rendering.

CCS Concepts: • **Computing methodologies** → **Perception**.

Additional Key Words and Phrases: contrast sensitivity, visual perception, foveated rendering

ACM Reference Format:
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 (July 2022), 16 pages. https://doi.org/10.1145/3528223.3530115

Authors' addresses: Rafał K. Mantiuk, rafal.mantiuk@cl.cam.ac.uk, University of Cambridge, William Gates Building, 15 JJ Thomson Avenue, Cambridge, UK, CB3 0FD; Maliha Ashraf, Maliha.Ashraf@liverpool.ac.uk, University of Liverpool, Eleanor Rathbone Building, Bedford Street South, Liverpool, UK, L69 7ZA; Alexandre Chapiro, alex@chapiro.net, Reality Labs, , Sunnyvale, USA, .

0730-0301/2022/7-ART145
https://doi.org/10.1145/3528223.3530115

1 INTRODUCTION

The spatial and temporal contrast sensitivity of the human visual system (HVS) has been studied by vision scientists for almost 70 years [De Lange 1952; Robson 1966; Schade 1956]. This form of measurement touches on a very fundamental aspect of vision, modeling the threshold response to luminance variation. The models of contrast sensitivity, known as contrast sensitivity functions (CSFs), explain the magnitude of contrast necessary to detect a pattern described by the parameters of the CSF. One attractive feature of the CSFs is that they are end-to-end models, which explain the response of the visual system (detection) for a given input (luminance pattern). Because of that, CSFs found many applications in image/video visibility and quality metrics [Andersson et al. 2020; Daly 1992; Mantiuk et al. 2005, 2021, 2011], compression codecs [Ahumada and Peterson 1992; Zeng et al. 2002], tone-mapping operators [Mantiuk et al. 2008], foveated rendering [Tursun et al. 2019] and many other areas of computer graphics.

The major challenge of measurement and modeling of the CSF is the large number of parameters describing the stimulus. All CSFs account for spatial frequency, but equally important is the background luminance and size of the stimulus. If we consider moving patterns, we also need to account for temporal frequencies. If the pattern is not fixated (i.e., is projected outside fovea), a CSF also needs to account for the position in the visual field, typically described by the eccentricity. Other relevant dimensions include the orientation of the (sinusoidal) pattern, the chromatic (color) coordinates of the background, and the chromatic direction of modulation [Wuerger et al. 2020]. In this work, we focus on the first 5 major dimensions of the CSF and consider only achromatic patterns. In contrast, most

existing CSFs account for only 3- or 4-dimensions of this space. For instance, traditional functions can model the visibility of moving patterns shown in the fovea, or static patterns shown in any part of the visual field, but not both.

The study of the CSF in 5 dimensions is made especially difficult as the entire space cannot be measured in a single experiment. This is due to the combinatorics of high-dimensional functions: consider that if we were to collect only 10 points along each dimension, and each measurement took 90 seconds (which is typical for the 30-50 trials this requires), we would need $10^5 \times 90/3600 = 2\,500$ h of experiment for a single observer. For that reason, most measurements of contrast sensitivity consider only a few slices from this multi-dimensional space.

To model a 5-dimensional CSF, we need to have good coverage of the parameter space. For that reason, we combine the data from 11 individual studies to build a standardized dataset that allows us to (a) choose the right visual models and relations; and (b) fit and validate the model. While previous works on CSF modeling attempted to fit one dataset at a time, our goal is to have a single model that can predict the data from all datasets without refitting. Combining and standardizing contrast sensitivity measurements is a non-trivial task as the measurements were often done using different protocols, detection criteria, viewing conditions (natural vs. artificial pupil, monocular vs. binocular, stabilized vs. unstabilized stimuli, etc.) and with differences in stimuli. The measurement from multiple sources can be used together only if the differences between the sources, for example in the stimulus size, can be accounted for by the CSF model. For that reason, using the data from multiple sources requires modeling all major parameters of the CSF.

The particular selection of the dimensions we model is motivated by the applications in augmented and virtual reality (AR/VR). We model both *spatial* and *temporal* contrast sensitivity, to allow for video applications. The effect of luminance is especially important given the modern display landscape, where available displays can reach luminance of 10 000 cd/m^2. Peripheral vision is of key importance to presentation on wide field-of-view displays, in particular in AR/VR. Finally, the model needs to account for the effect of size to conciliate the difference between datasets we use for modeling.

In summary, our contributions are as follows:

- We created a large dataset of contrast sensitivity data, combining and standardizing 11 published studies, which we make publicly available.
- We perform (previously impossible) quantitative analyses of existing CSF models, demonstrating their ability to predict the measured data.
- We introduce stelaCSF, a new model encompassing 5 major parameters of contrast sensitivity, and outperforming existing functions by a large margin. The code of our model is publicly available[1].
- We demonstrate sample applications of our CSF model to relevant problems in computer graphics, improving methods for flicker detection and foveated rendering.

[1]Code and data available at the project web page: https://www.cl.cam.ac.uk/research/rainbow/projects/stelaCSF and https://github.com/gfxdisp/stelaCSF.

2 RELATED WORK

The method of measuring the eye's response to sinusoidal gratings to measure the performance of the human visual system was first introduced by Schade [1956]. The work models the human visual system as an analog sequential system with a single channel bound by the limitations of the optical elements of the eye as well as the post-retinal systems. Later work by Campbell and Robson [1968] presented the idea of the visual system as a collection of narrow-band spatially-tuned channels, analogous to a Fourier decomposition. The contrast sensitivity function (CSF) was thus hypothesized to be the envelope formed by the combined response of these individual spatial channels.

The attempts to model the CSF are too numerous to list in this short review. We focus on a few representative examples, in particular those that are commonly used in graphics or are relevant for our model.

In a series of papers, Rovamo and colleagues measured contrast sensitivity across sizes [Rovamo et al. 1993], luminance levels [Mustonen et al. 1993], eccentricities [Rovamo et al. 1995; Virsu and Rovamo 1979], and proposed simple relationships governing the sensitivity across those dimensions. Because our own modeling relies on the data and ideas from these papers, we combined the proposed components and use them as a baseline model. A Visual Difference Predictor, proposed by Daly [1992], relies on one of the most complete models of contrast sensitivity, though details on its creation are scarce. Barten took a principled approach to modeling different sources of noise and attenuation found in the visual system and proposed both general [Barten 1999] and simplified CSF formulas [Barten 2003], which have been widely adopted in various applications and standards, such as SMPTE 2084 and ITU-R 2100. Although the formula has been derived for photopic (cone-mediated, daylight) vision, it has also been shown to predict mesopic and scotopic data [Barten 2003]. Both Barten [2003] and Yi et al. [2022] add the dependency of the contrast sensitivity on surround luminance. Here, we assume that the effect of surround luminance can be modeled separately as glare and local adaptation [Vangorp et al. 2015].

All the CSFs mentioned so far model the sensitivity as a function of spatial frequency, size, luminance and eccentricity (except Barten's model, which lacks eccentricity), but they do not account for temporal frequencies. Kelly conducted a systematic study of spatio-temporal patterns and proposed a spatio-temporal CSF [1979b], which was later adjusted by Daly [1998] to better describe naturally observed stimuli shown on a CRT display. This is because Kelly's measurements were collected for stimuli that compensated for eye movements, which strongly shifted the contrast sensitivity towards higher frequencies [Kelly 1979a]. Kelly-Daly's model, however, does not account for any other dimensions of the CSF.

Watson and Ahumada made the observation that the logarithm of contrast sensitivity can be well explained as a linear model of spatial and temporal frequency, and logarithmic retinal illuminance (luminance multiplied by the area of the pupil) [Watson and Ahumada 2016]. They named their proposed simplified model of the CSF the pyramid of visibility (PoV). Although the PoV is very attractive due to its simplicity, it suffers from one major limitation — it cannot

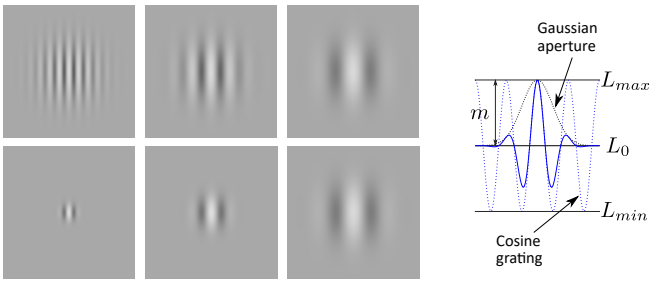

Fig. 2. Gabor patches of constant size and decreasing frequency (top left). Gabor patches of constant number of cycles and decreasing frequency (bottom left). The luminance profile of the right-most patch (right).

model the frequencies under 10 cpd and 10 Hz. This is the frequency range that is most visible to the human eye and is a fundamental part of visual content (for instance, most commercially available VR headsets can only reproduce values up to 8 cpd). The PoV model was later extended to account for the effect of eccentricity [Watson 2018].

Ahumada et al. proposed to model spatio-temporal sensitivity using two temporal channels, motivated by the existence of parvo- and magnocellular pathways in the visual system [2018]. We take a similar approach to modeling temporal sensitivity and demonstrate that a two-channel model can predict a 5-dimensional space of contrast sensitivity well.

Complementary to the CSF data are the measurements of the critical flicker fusion frequency (CFF). These measurements capture the temporal frequency at which sensitivity drops to the boundary of visibility (as we will explain in Section 3). Krajancich et al. measured the CFF for eccentricities up to 55 deg, a range of low spatial frequencies and luminance levels [2021]. They then fit a parametric model to their data. Although the CFF has many uses, it defines only the boundary of visibility, and does not explain contrast sensitivity inside that boundary.

The only model we are aware of which accounts for the same 5 dimensions as stelaCSF, is the contrast sensitivity used in the FovVideoVDP video quality metric [Mantiuk et al. 2021]. Their model is a combination of the spatio-chromatic CSF from [Mantiuk et al. 2020], the cortical magnification model, and Kelly-Daly's model, discussed above. We show in Section 6.2 that this combination introduces a large prediction error for both low and high spatial frequencies.

The contrast sensitivity models discussed above were often shown to provide excellent fits to data. However, this performance happened when fitting to one dataset at a time. In contrast to previous studies, we do not attempt to obtain a perfect fit to individual datasets, but instead have a single model that can explain all datasets without refitting.

3 CONTRAST DETECTION

We now explain the notation used throughout the rest of the paper and introduce the fundamental concepts of contrast sensitivity. This section is meant to provide an introduction to contrast sensitivity and its measurements to those who are less familiar with this topic.

The majority of the CSF data comes from experiments with cosine (or sine) gratings, modulated in the spatial or both spatial and temporal dimensions:

$$s(x,t) = L_0 + m\,\cos 2\pi\rho x\,\cos 2\pi\omega t \quad [\,\mathrm{cd/m^2}\,], \tag{1}$$

where x represents spatial horizontal coordinates in visual degrees, t is time in seconds, L_0 is the luminance of the background in $\mathrm{cd/m^2}$, ρ is the *spatial frequency* in cycles per degree (cpd), ω is the *temporal frequency* in Hz and m is the *modulation* of the grating. Some authors use a relative modulation $m_{\mathrm{rel}} = m/L_0$. Here, we will use a regular modulation to distinguish it from contrast, introduced later. The grating does not need to be horizontal, as one in Eq. (1), and the equation can be easily modified to generate grating of any orientation. To restrict the size of a grating, the stimulus is typically limited by a square or circular aperture, with either a smooth or sharp boundary. More recent measurements use a Gaussian function to restrict the aperture:

$$s'(x,y,t) = s(x,t)\,\exp\left(-\frac{x^2+y^2}{2\sigma^2}\right) \quad [\,\mathrm{cd/m^2}], \tag{2}$$

where σ controls the size of the aperture. If the Gaussian function above is used, the stimulus is known as a Gabor patch. An example of such a patch is shown in Figure 2. If the stimulus size is adjusted to show the same number of λ-cycles,

$$\sigma = \frac{\lambda}{2\rho} \quad [\mathrm{deg}]\,, \tag{3}$$

the stimulus is described as having constant cycles. If the size of the stimulus is constant regardless of the spatial frequency, the stimulus is described as being constant size. A similar exposure window is also often introduced in the temporal dimension to restrict the duration of a stimulus.

Contrast detection data for gratings is typically reported in units of Michelson contrast:

$$c = \frac{L_{\mathrm{max}} - L_{\mathrm{min}}}{L_{\mathrm{max}} + L_{\mathrm{min}}} = \frac{m}{L_0}\,, \tag{4}$$

where L_{max} and L_{min} are the maximum and minimum luminance of the stimulus, as shown in Figure 2. For sine/cosine gratings, Michelson contrast is equivalent to the modulation amplitude m divided by the background luminance L_0 (both from Eq. (1)). The maximum contrast value is 1, as a contrast above 1 would result in negative luminance values in Eq. (1). Many papers report threshold values in units of sensitivity, which is the inverse of detection contrast:

$$S = \frac{1}{c} = \frac{L_0}{m}\,. \tag{5}$$

Because the maximum contrast is 1, the minimum sensitivity value is also 1.

The goal of contrast detection experiments is to find the smallest contrast c, at which the stimulus is detected in p_{thr} percent of the cases, where p_{thr} varies from 0.75 to 0.83, depending on the experimental procedure. This percentage is typically measured using an *n-alternative-force-choice* (2AFC, 3AFC,...) protocol, in which n stimuli are shown sequentially or side-by-side and the observer needs to select the one that contained the stimulus or showed a different orientation from the rest. Such experiments require that the stimulus is measured multiple times at the same and at different

Table 1. The datasets used in our study. A (∗) means that we obtained the data directly from the authors. In the remaining cases, the data were scanned from plots in the relevant publication.

Dataset Name	Spatial freq. cpd	Temporal freq. Hz	Luminance cd/m^2	Eccentricity deg (visual field)	Area deg^2	Stimulus	Criterion
Modelfest [Watson 2000] (∗)	1.12 - 30	✗	30	0	0.003 - 0.78	Gabor patch	Contrast detection
HDR-VDP CSF [Mantiuk et al. 2011] (∗)	0.125 - 32	✗	0.002, 0.02, 0.2, 2, 20, 150	0	0.07, 0.78, 7.06	Fixed cycles Gabor patch	Contrast detection
HDR CSF [Wuerger et al. 2020] (∗)	0.125, 0.25, 0.5, 1, 2, 4, 6, 12, 24	✗	0.002, 0.02, 0.2, 2, 200, 2000, 10000	0	0.05 - 50	Fixed cycles Gabor patch	Contrast detection
Rovamo et al. [1993]	0.125, 0.25, 0.5, 1, 2, 4, 8, 16, 32	✗	50	0	0.003 - 980	Grating with rectangular aperture	Contrast detection
Robson [1966]	0.5 - 30	0.5 - 32	20	0	6.25	Grating with rectangular aperture	Contrast detection
Laird et al. [2006]	4, 8, 16	9.2 - 31.4	60	0	4.75	Gabor patch	Contrast detection
Snowden et al. [1995]	0.25, 1, 2, 4, 5, 10, 20	0.8 - 55.7	0.02 - 870	0	0.25, 1, 4.01	Gabor patch	Contrast detection
Virsu and Rovamo [1979]	0.5 - 16	✗	10	0, 5, 10, 15, 20, 25, 30 (nasal)	19.65	Grating with circular aperture	Contrast detection
Virsu et al. [1982]	1 - 22.6	1, 18	10	0, 1.5, 4, 7.5, 14, 30 (nasal)	1.57	Grating with semi-circular aperture	Contrast detection
Wright and Johnston [1983]	0.25, 2, 6, 9	0, 0.25, 8, 16	100	0 - 12 (superior)	0.75, 2.35, 36	Grating with rectangular aperture	Contrast detection
Anderson et al. [1991]	0.15 - 10	8	225	8, 25, 40, 55 (nasal, temporal, inferior, superior)	0.78, 19.6	Gabor patch	Direction of motion discrimination

contrasts c, so that a smooth function could be fitted to the collected observations — i.e. to the probabilities of providing a correct answer at a given contrast c. Such a fitted function, known as a *psychometric function*, is used to estimate the contrast c at which probability reaches p_{thr}. A faster, but less precise protocol may involve a *method of adjustment*, in which an observer directly controls c and confirms their choice once the stimulus starts to become visible or invisible. Such a protocol typically results in higher variance, as it measures subjective judgments (i.e. whether the observer thinks the stimulus has become invisible) rather than objective performance (can the observer detect the stimulus). Another problem is that the threshold can be influenced by the direction in which the adjustment is made — from visible to invisible contrast or vice versa. To alleviate this, the threshold is often measured multiple times in each direction.

It must be stressed that the stimuli are always shown on calibrated monitors of sufficient bit-depth (typically 10-12 bits), that can reproduce the linear units of luminance. At very high spatial frequencies, the linear grating from Eq. (2) will be attenuated by the

optics of the eye (low-pass filtered), and the participant will see a flat field of luminance L_0. If the same grating is shown on a non-calibrated monitor, the observer will notice the bias in luminance in the shape of the Gaussian aperture, resulting in incorrect detection thresholds.

The spatial or temporal frequency at which the sensitivity $S = 1$ (or sometimes $S = 2$) is assumed to be the boundary of visibility. It marks the highest frequency that can be seen by the eye. For temporal frequencies, that point is known as the *critical flicker frequency* (CFF) [Hartmann et al. 1979; Tyler and Hamer 1990, 1993].

It is also important to stress that the CSF is meant to predict only low-contrast, just-visible (near-threshold) differences, shown on a uniform background. The CSF alone does not explain well-visible (supra-threshold) differences or the visibility of patterns shown on patterned backgrounds as it does not account for contrast constancy [Georgeson and Sullivan 1975] nor for contrast masking [Legge and Foley 1980]. However, it is a building block of models that account for those effects [Daly 1992; Mantiuk et al. 2011]. CSF also cannot predict the effect of crowding, which results in lower rate of identification when a pattern in periphery is surrounded by other patterns [Pelli et al. 2004]. The CSF is not a linear model of the visual system as even early vision is highly non-linear and cannot be approximated using linear band-pass or low-pass filters (with the exception of high spatial and temporal frequencies).

4 DATASETS

The 11 datasets that we used to model, fit and validate CSFs, are listed in Table 1. These were selected to (a) provide good coverage of interactions in the 5-dimensional space of contrast sensitivity; (b) were measured for Gabor patches and cosine gratings, which can be modeled by a CSF. For some of the datasets, we were able to obtain the data directly from the authors. For the rest, the data was collected by scanning points from the plots in the corresponding papers using WebPlotDigitizer software [Rohatgi 2021] . Some of the experiments were performed by 20–25 observers (HDR-CSF), while others were collected only by the study's author (Robson). We did not distinguish between these cases, and fitted the models to the average observer data. Our fitting procedure accounted for individual differences, as we will explain in Section 6.

Spatial frequency is typically well represented and the data spans the range from 0.125 to 32 cpd across all datasets. 6 of the datasets were measured for a range of *temporal frequencies*, while static stimuli were used for the rest. We assumed that all static stimuli had the temporal frequency of 0 Hz (we ignore the drift due to the movement of a fixated eye). Most datasets were measured for the photopic *luminance* levels between 10 and 100 cd/m^2, but we also included the datasets with scotopic luminance down to 0.002 cd/m^2 and one dataset that measured contrast sensitivity up to 10 000 cd/m^2. For all dataset that controlled *eccentricity*, we registered the position in the visual field (nasal, temporal, superior, inferior) to account for the anisotropy of the extrafoveal vision. When a stimulus is presented at larger eccentricities on a flat display and observer's visual axis is perpendicular to the display, the effective spatial frequency is increased because of foreshortening (see. [Mantiuk et al. 2021, 3.1]). The studies we rely on avoided this issue by tilting the display so that it was facing the eye [Virsu and Rovamo 1979, Fig.1] removing the need for additional adjustments.Some datasets relied on constant-cycles stimuli (HDR-CSF) but most used constant size stimuli. When the dataset used a smooth Gaussian aperture, we assumed that the area of the stimulus corresponded to the area of a disk with a radius equal to the standard deviation of the Gaussian aperture.

5 MODEL

Our goal is to find a general model that accounts for all major dimensions of achromatic contrast sensitivity: spatial and temporal frequency, luminance, size and eccentricity. The predictions should be accurate for the entire visible range of each dimension, including low spatial and temporal frequencies down to 0.1 cpd and 0 Hz, and the full visible range of luminance, from scotopic to the bright photopic light levels. To ensure that the model can generalize, we want to predict the data from all 11 datasets using the same parametric fit of the model. We do not expect to fit all datasets perfectly as the measurement conditions were often too different. However, the fits should be plausible. Whenever possible, we base our model on existing psychophysical models of contrast sensitivity, or support the choice with relevant research. If a certain aspect of contrast sensitivity is not well represented in any of the datasets, we opt for a simpler model.

5.1 Size and spatial frequency

Larger patterns are more likely to be detected. Rovamo et al. noted that the sensitivity to vertical gratings (square aperture) is proportional to the square root the grating area multiplied by the spatial frequency [Rovamo et al. 1993]. For square-shaped gratings, the sensitivity happens to be proportional to the number of cycles in the grating:

$$S \propto \sqrt{a}\,\rho = d\,\rho \quad [\text{cyc}]\,, \tag{6}$$

where a is the area of the stimulus in deg^2, ρ is the spatial frequency in cycles-per-degree (cpd), and d is the length of the side of the square aperture in degrees. However, the sensitivity increases only to a certain critical area, a_c, which Rovamo et al. modeled as:

$$a_\text{c} = \frac{a_0}{1 + (\rho/\rho_0)^2} \quad [\text{deg}]\,, \tag{7}$$

where $a_0 = 270\,\text{deg}$ and $\rho_0 = 0.65\,\text{cpd}$. This quantity is easier to interpret as the critical number of cycles ($\sqrt{a_c}\,\rho$) over which we can integrate, which is plotted in the top-left of Figure 3.

Rovamo et al. propose to model area-dependent contrast sensitivity as:

$$S_\text{R}(\rho, a) = S_\text{m} \sqrt{\frac{a_\text{c}}{1 + a_\text{c}/a}}\,\rho\,OTF(\rho)\,, \tag{8}$$

where S_m is the peak sensitivity, $OTF(\rho)$ is the optical transfer function of the eye, modeled as a low-pass filter. The strength of this model is that it not only accounts for spatial integration, but also for the drop of sensitivity at low spatial frequencies, as shown in the bottom-left of Figure 3. It is well known that the sensitivity strongly depends on the number of displayed cycles at low spatial frequencies [Savoy and McCann 1975] and the model accounts for that (see Figure 3 bottom-left, dashed lines). However, the model also has two weaknesses: contrary to the data from [Wuerger et al.

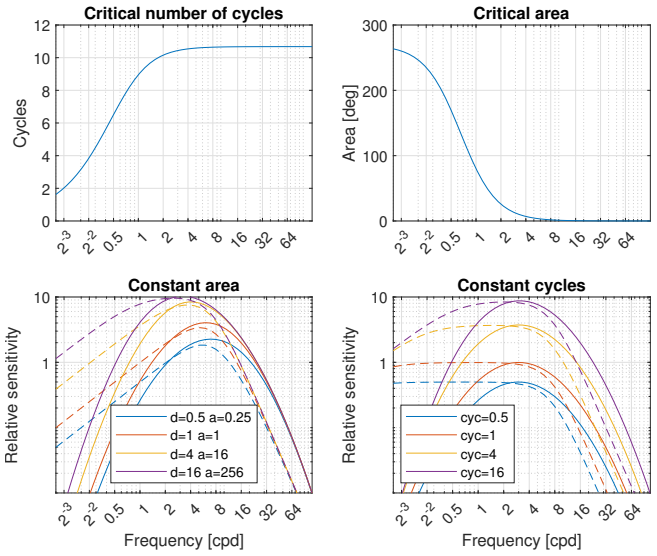

Fig. 3. Top: the critical number of cycles (left) and area (right) that limits the extent of spatial integration. Bottom: The relative sensitivity for the patterns of fixed area (left) or the fixed number of cycles (right), as predicted by Rovamo et al. model (dashed lines) and our model (continuous lines).

2020], it shows no drop of sensitivity at low frequencies for constant-cycle Gabors (see Figure 3 bottom-right, dashed lines, and also Figure 10); and the OTF cannot account for the loss of sensitivity at high frequencies outside the fovea [Anderson et al. 1991, Fig. 5]. Because of that, we modify the original model so that it relies on the truncated log-parabola [Ahumada and Peterson 1992]:

$$S_{\mathrm{A}}(\rho, a, L) = S_{\mathrm{m}}(L)\sqrt{\frac{a_{\mathrm{c}}}{1 + a_{\mathrm{c}}/a}}\,\rho\, S_{\mathrm{LP}}(\rho)\,, \tag{9}$$

where the truncated log-parabola is:

$$S_{\mathrm{LP}}(\rho) = \begin{cases} 1 - k_{\mathrm{a}} & \text{if} \quad \rho < \rho_{\mathrm{m}} \quad \text{and} \quad S_{\mathrm{LP}} < 1 - k_{\mathrm{a}} \\ 10^{-\frac{(\log_{10}\rho - \log_{10}\rho_{\mathrm{m}}(L))^2}{2k_{\mathrm{b}}}} & \text{otherwise} \end{cases} \tag{10}$$

ρ_{m} controls the position of the peak of the parabola as a function of luminance and will be introduced in Section 5.3. k_{b} controls its bandwidth, and k_{a} is responsible for limiting the drop at low frequencies. The sensitivity functions for patterns of different areas and different numbers of cycles are shown in the bottom part of Figure 3.

Because spatial integration does not depend on luminance [Mustonen et al. 1993], nor on eccentricity [Rovamo et al. 1995] we do not need to account for the interaction of those parameters.

5.2 Temporal frequency

It is argued that visual information is processed by two temporal channels: one responsible for encoding of low temporal frequencies and another for high temporal frequencies. When researchers argue that this separation stems from the retina, these channels are known as *sustained* and *transient* channels respectively [Burbeck and Kelly 1980; Hammett and Smith 1992]. If instead researchers point to the

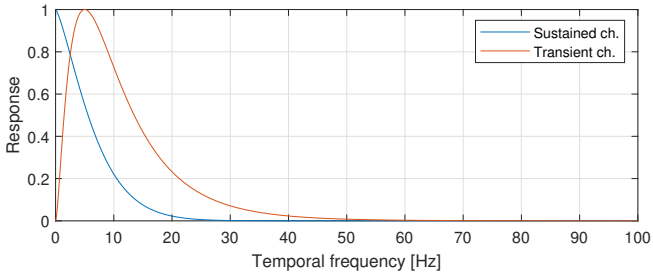

Fig. 4. The responses of the two temporal channels, tuned to low and high temporal frequencies.

LGN or cortex, the channels are called *parvocellular* and *magnocellular* channels [Merigan et al. 1991]. Without making arguments about the site of temporal channel separation, we will denote the low frequency channel by the subscript $\mathcal{S}$ and the high frequency channel by $\mathcal{T}$. The response of temporal channels is often expressed by a cascade of exponential filters, which introduces complexity into modeling. We found that the steady-state responses can be well-approximated by the generalized exponential functions:

$$R_{\mathcal{S}}(\omega) = \exp\left(-\frac{\omega^{\beta_{\mathcal{S}}}}{\sigma_{\mathcal{S}}}\right) \tag{11}$$

and:

$$R_{\mathcal{T}}(\omega) = \exp\left(-\frac{\left|\omega^{\beta_{\mathcal{T}}} - \omega_0^{\beta_{\mathcal{T}}}\right|^2}{\sigma_{\mathcal{T}}}\right), \tag{12}$$

where ω is the temporal frequency in Hz, $\beta_{\mathcal{S}}$, $\sigma_{\mathcal{S}}$, $\beta_{\mathcal{T}}$ and $\sigma_{\mathcal{T}}$ are the parameters of the model. The responses of the temporal channels are plotted in Figure 4.

5.3 Luminance

In dim light, contrast sensitivity increases in proportion to the square root of retinal illuminance, according to the DeVries-Rose law, but in bright light contrast sensitivity follows Weber's law and is independent of illuminance [Rovamo et al. 1995]. The recent sensitivity data from [Wuerger et al. 2020], measured up to 10 000 cd/m^2, shows that this statement is not entirely accurate as the sensitivity at very high luminance, above 1000 cd/m^2, starts to drop. Luminance also causes the shift of the CSF towards lower frequencies, as the reduction of sensitivity with luminance is stronger for high frequencies. Those effects have different characteristics for sustained and transient temporal channels. We model both effects of luminance separately for the two temporal channels as:

$$\begin{aligned} S_{\mathrm{m},\mathcal{S}}(L) &= k_{\mathrm{s1},\mathcal{S}}\left(\frac{1 + k_{\mathrm{s2},\mathcal{S}}}{L}\right)^{k_{\mathrm{s3},\mathcal{S}}}\left(1 - \frac{1 + k_{\mathrm{s4},\mathcal{S}}}{L}\right)^{k_{\mathrm{s5},\mathcal{S}}} \\ S_{\mathrm{m},\mathcal{T}}(L) &= k_{\mathrm{s2},\mathcal{T}} L^{k_{\mathrm{s1},\mathcal{T}}} \end{aligned} \tag{13}$$

and the effect on the peak frequency shift as:

$$\begin{aligned} \rho_{\mathrm{m},\mathcal{S}}(L) &= k_{\rho 1,\mathcal{S}}\left(\frac{1 + k_{\rho 2,\mathcal{S}}}{L}\right)^{k_{\rho 3,\mathcal{S}}} \\ \rho_{\mathrm{m},\mathcal{T}}(L) &= k_{\rho,\mathcal{T}} \end{aligned} \tag{14}$$

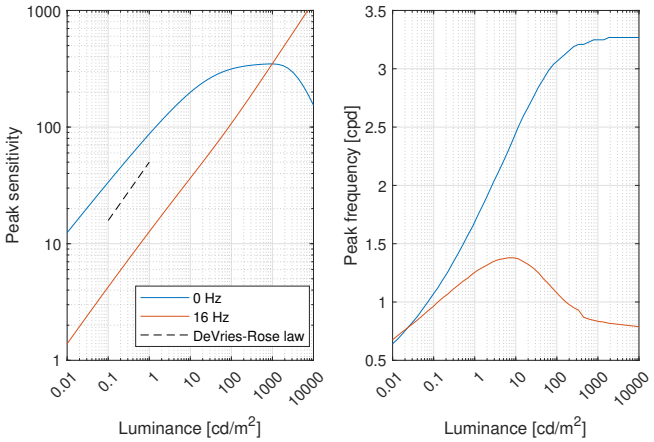

Fig. 5. The peak sensitivity (left) and frequency (right) for the patterns of 0 Hz (blue) and 16 Hz (red). The grating is shown with a disk aperture of the radius 1.5 deg. The dashed line shows the slope corresponding to the DeVries-Rose law.

where L is luminance in cd/m^2, $k_{...}$ are the parameters of the model, $S_{\mathrm{m},\mathcal{S}/\mathcal{T}}$ is the luminance-dependent sensitivity from Eq. (9), and $\rho_{\mathrm{m},\mathcal{S}/\mathcal{T}}$ is responsible for the shift of peak frequency in Eq. (10).

It should be noted that those functions do not define the location of the peak of the CSF, as the peak is affected by the area-dependent component from Eq. (9). The peaks of our model are found numerically, and plotted in Figure 5. This plot shows that the sensitivity of the low-frequency channel ($\mathcal{S}$) increases approximately according to the DeVries-Rose law at low luminances, then smoothly transitions to a short Weber region (flat), and drops at high luminance levels. The high-frequency channel ($\mathcal{T}$) follows the DeVries-Rose law in the entire range of luminance as we did not find data showing the transition to the Weber region. The peak frequency strongly depends on the size of the stimulus and can vary between 0.1 and 3 cpd.

5.4 Eccentricity

Contrast sensitivity is highest in the fovea and decreases with eccentricity (distance from the center of fovea in visual degrees), most likely due to the reduced density and larger pooling area of the ganglion cells [Anderson et al. 1991]. This effect is typically explained by the cortical magnification [Strasburger et al. 2011]. It relies on the observation that the visibility of patterns in the fovea and periphery can be matched by enlarging the pattern by a factor that depends on eccentricity (the inverse of the cortical magnification) [Virsu and Rovamo 1979]. If we want to account for cortical magnification, we need to reduce the size of the stimulus and increase its spatial frequency by a factor given by the cortical magnification (see for example [Mantiuk et al. 2021, Sec. 3.6]).

We have experimented with multiple variants of the cortical magnification models and found them to be inconsistent with several datasets. The change of temporal frequency by the cortical magnification factor results in a very strong shift of the peak of the CSF towards lower frequencies, which cannot be observed in the data plotted in Figure 15 nor in Figure 16. The dashed lines in those plots indicate the model that employs cortical magnification but fails to predict the data. For this reason, we employed a simpler model that relies on the linear relation between log-sensitivity, eccentricity and the cycles of the underlying frequency [Watson 2018].

We model the drop of sensitivity with eccentricity e in the same way as [Watson 2018]:

$$S_{\mathrm{ecc}}(e,\rho) = 10^{\hat{k_{\mathrm{e1}}}\rho e + \hat{k_{\mathrm{e2}}} e} \tag{15}$$

The effect of eccentricity and spatial frequency on sensitivity is shown in Figure 6, 1st row and 4th column. The original model does not account for the anisotropy of peripheral vision; the reduction of sensitivity is slower in the nasal than in the temporal direction [Anderson et al. 1991]. We account for that by fitting a separate set of parameters for the nasal and other directions in the visual field. We find those parameters separately and interpolate between the orientations in the visual field:

$$\begin{aligned} &\hat{k_{ei}} = \alpha k_{ei} + (1-\alpha) k_{ei,\mathrm{nasal}} \quad \text{where} \quad i = 1,2 \quad \text{and} \\ &\alpha = \min\left\{1, \left|\frac{\theta - 180}{90}\right|\right\} \end{aligned} \tag{16}$$

θ is the orientation in the visual field in deg. $\theta = 0$ corresponds to the temporal and $\theta = 180$ to the nasal directions; i.e. it is an angular coordinate for the right eye.

5.5 Final model

The final value of contrast sensitivity in the 5-dimensional space is predicted by combining the components introduced in the previous sections:

$$\begin{aligned} &S(\rho,\omega,L,a,e) = \\ &S_{\mathrm{ecc}}(e,\rho)\left(R_{\mathcal{S}}(\omega)\, S_{\mathrm{A},\mathcal{S}}(\rho,a,L) + R_{\mathcal{T}}(\omega)\, S_{\mathrm{A},\mathcal{T}}(\rho,a,L)\right) \end{aligned} \tag{17}$$

The two temporal channels, with the temporal frequency responses $R_{\mathcal{S}}$ and $R_{\mathcal{T}}$, are independent and have different sensitivity characteristic ($S_{\mathrm{A},\mathcal{S}/\mathcal{T}}$). The model assumes the same drop of sensitivity with eccentricity, S_{ecc}, which is independent from stimulus size, luminance and temporal frequency. This assumption is a necessary simplification, however, it results in worse predictions at high temporal frequencies, as we will explain in Section 6.3.

6 MODEL FITTING

In this section we compare how well stelaCSF and existing models can predict the data from multiple datasets. For fair comparison, we optimize the free parameters of each model. Below we explain the optimization procedure.

To be able to fit contrast sensitivity models to multiple datasets, we need to introduce a per-dataset adjustment factor, s_d. It is a gain control parameter which shifts the CSF up or down in log-space and does not influence the shape of the function. This factor is necessary to account for all the differences in stimuli and the experimental procedure that are not explained by the parameters of the CSF. It also accounts for the individual differences, as some datasets were collected for just one observer. Ideally, we want this factor to be close to 1. Therefore, our fitting loss function is:

$$\mathcal{L} = \frac{1}{N}\sum_d \sum_i \left(\log_{10} S_{i,d} - s_d \log_{10} \tilde{S}_{i,d}\right)^2 + \frac{\lambda}{D}\sum_d (\log_{10} s_d)^2 \tag{18}$$

Table 2. The fitted parameters of stelaCSF.

Part	Parameters
Size and spatial freq.	$k_{a,\mathcal{S}} = 0.0835$, $k_{b,\mathcal{S}} = 1.126$, $k_{a,\mathcal{T}} = 0.00027$, $k_{b,\mathcal{T}} = 2.83$
Temporal channels	$\beta_{\mathcal{S}} = 1.331$, $\sigma_{\mathcal{S}} = 0.1319$, $\beta_{\mathcal{T}} = 0.1898$, $\sigma_{\mathcal{T}} = 0.1319$,
Luminance	$k_{s1,\mathcal{S}} = 65.4$, $k_{s2,\mathcal{S}} = 29.83$, $k_{s3,\mathcal{S}} = 0.2038$, $k_{s4,\mathcal{S}} = 7.634e-7$, $k_{s4,\mathcal{S}} = 6.936e9$, $k_{\rho 1,\mathcal{S}} = 1.565$, $k_{\rho 2,\mathcal{S}} = 51.19$, $k_{\rho 3,\mathcal{S}} = 0.245$, $k_{s1,\mathcal{T}} = 0.5278$ $k_{s2,\mathcal{T}} = 64.06$ $k_{\rho,\mathcal{T}} = 0.01629$,
Eccentricity	$k_{e1} = 0.0201$, $k_{e2} = 0.02694$, $k_{e1,\text{nasal}} = 0.01351$, $k_{e2,\text{nasal}} = 0.01713$,

where $d = 1, \ldots, D$ is the index of the dataset, i is the index of a stimulus in the dataset, N is the total number of stimuli, $S_{i,d}$ and $\tilde{S}_{i,d}$ are the reference and predicted sensitivity values. We set λ to 0.01. In all our experiments, we fix $s_d = 1$ for the ModelFest dataset. We fit all models using a quasi-Newton method implemented in Matlab's *fminunc* function.

The parameters of the fitted stelaCSF are listed in Table 2. Figure 6 visualizes all 5 dimensions of our model by plotting slices across two selected dimensions. It shows how the CSF changes its shape between low and high-pass depending on all other parameters (1st column); that the visibility of higher temporal frequencies depends mostly on luminance (2nd column and 2nd row); that the effect of eccentricity is predicted as a linear function of eccentricity (4th column); and that size depends only on spatial frequency as the vertical shift can account for all other dimensions (5th column).

The CSF can be used to determine the boundary of visibility — the surface at which the detection contrast is maximum (1) and therefore sensitivity is also 1. We numerically determined the boundary of visibility, and plotted it in Figure 7. The top two plots compare the pyramid of visibility (PoV) [Watson and Ahumada 2016] with stelaCSF, showing the missing apex with low spatio-temporal frequencies for PoV and a more intricate shape for stelaCSF. Because we model all 5 dimensions, we can find the boundary of visibility as the function of eccentricity and size for stelaCSF (the two bottom plots).

6.1 Model comparison

One challenge of comparing CSF models is that contrast sensitivity data cannot be easily split into datasets used for testing and training. This is because each dataset typically contains uniformly spaced samples across a few select slices of the 5$^{\text{th}}$ dimensional space. If a model is tested on a few slices of that space, the measure of error is not representative for the entire space. For that reason, we perform 5-fold cross-validation within each dataset (leave-one-out, 5 splits) and use all datasets for both training and testing. We also follow the approach from other works [Ahumada et al. 2018; Watson and Ahumada 2005] and report the result for the entire dataset without train/test split. The error is reported as root-mean-square-error (RMSE), represented in dB units:

$$\mathcal{E} = 20 \sqrt{\sum_d \sum_i \left(\log_{10} S_{i,d} - s_d \log_{10} \tilde{S}_{i,d} \right)^2} \quad \text{[dB]} \qquad (19)$$

Table 3. RMSE prediction error of contrast sensitivity models in dB. The columns correspond to the set of dimensions (and datasets) that was used with each model. The large font numbers correspond to the result for the entire dataset, while smaller font numbers denote the results of 5-fold cross-validation, representing the mean and standard deviation across all folds. Mark "✗" means that the model does not support the given set of dimensions.

Model	ρ, L, a	ρ, L, e, a	ρ, ω, L, e, a
Barten's CSF	4.58 4.47±0.70	✗	✗
VDP CSF	6.88 10.02±2.76	8.76 11.34±0.55	✗
Rovamo's CSF	4.23 6.15±0.75	5.63 8.33±0.78	✗
fvvdpCSF	3.71 4.52±0.78	6.36 7.55±0.46	7.92 10.55±1.10
stelaCSF	3.11 4.47±0.70	4.09 5.19±0.51	4.16 5.91±0.25

Even though the popular CSF models, such as that of Barten [Barten 2003], do not account for all five dimensions of contrast sensitivity, it is interesting to see how they perform compared to our model. For that purpose, we separately compared 3-, 4- and 5-dimensional CSFs. The 3-dimensional CSFs are all those that account for spatial frequency (ρ), luminance (L) and area (a). The 4-th dimensional CSFs add eccentricity (e) and 5-th dimensional also include the temporal frequency (ω). For the test with 3- and 4-dimensional CSFs, we used a subset of datasets that excluded the missing dimensions.

We included in the comparison Barten's [2003] and the CSF from VDP paper [Daly 1992], as they are widely used. We also implemented the model proposed by Rovamo et al. [1995] and combined it with the models of spatial integration [Rovamo et al. 1993], and cortical magnification [Virsu and Rovamo 1979]. The only existing CSF model that accounts for all 5 dimensions is the one proposed in the FovVideoVDP paper [Mantiuk et al. 2021], which we will abbreviate as fvvdpCSF. We found it impractical to compare stelaCSF with the pyramid of visibility (PoV) [Watson 2018; Watson and Ahumada 2016] because (a) the PoV does not account for the effect of stimulus size resulting in large prediction errors; and (b) the PoV cannot provide valid predictions when both spatial and temporal frequencies are less than 10 deg and 10 Hz respectively, which restricts the data it can be tested on.

For a fair comparison, we fitted the parameters of each model separately for the 3-, 4- and 5-dimensional CSF comparison. To decide on the set of parameters that should be fitted, we tested whether including a parameter improved the fit and did not result in implausible predictions, such as a sharp rise in sensitivity at high frequencies. The latter test was performed by inspecting the plots similar to the ones shown in Figure 6.

The overall performance for all models is shown in Table 3. The results indicate that stelaCSF can explain the datasets better than existing models, regardless of the number of considered dimensions. The differences in the fitting errors between stelaCSF and fvvdpCSF

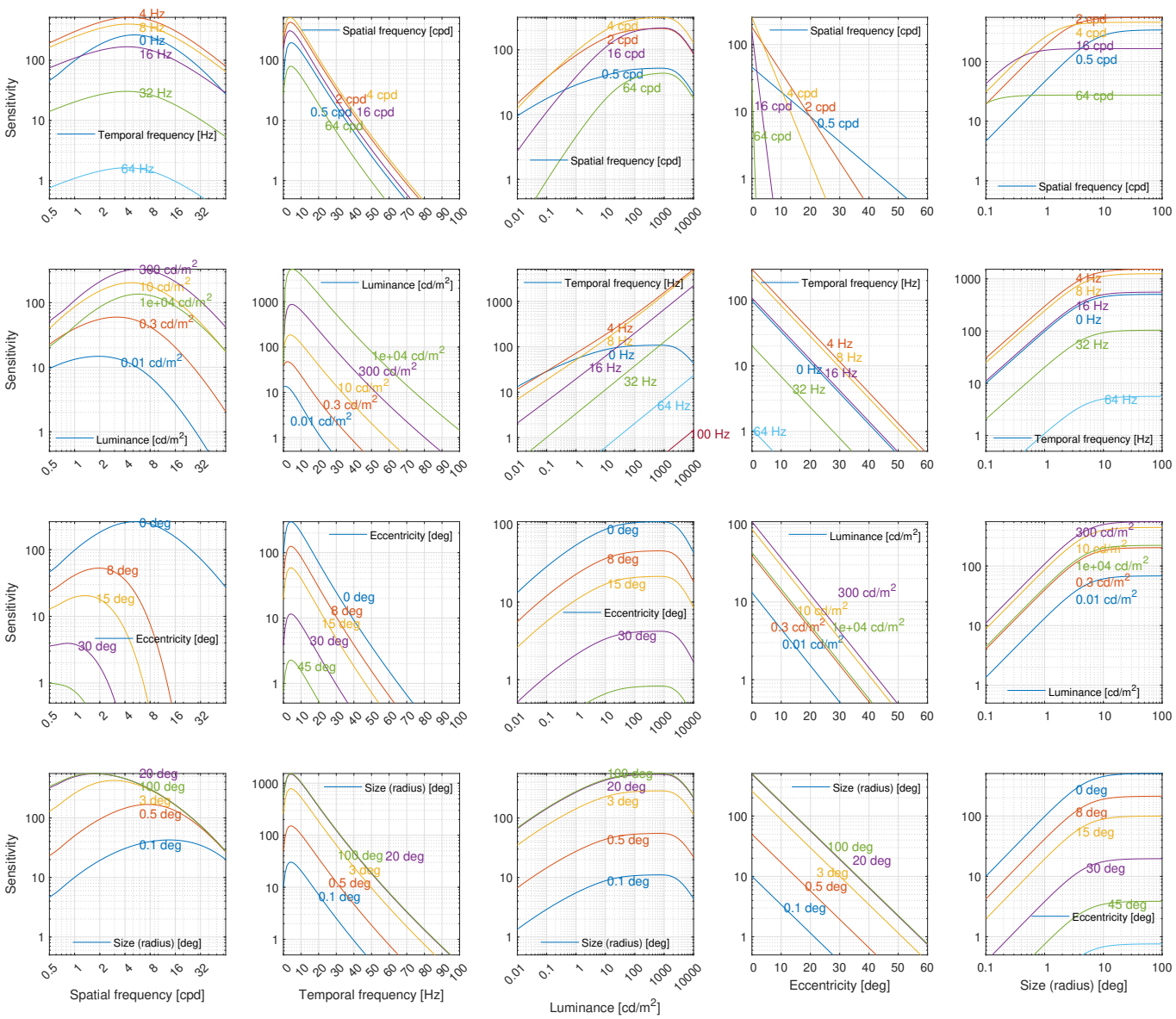

Fig. 6. Five dimensions of stelaCSF visualized for all combinations of pairs of parameters. The other parameters were set as follows: $\rho = 1$ [cpd], $\omega = 0$ [Hz], $L = 30$ [cd/m^2], $a = 3.14$ [deg^2], $e = 0$ [deg] (temporal eccentricity).

were compared using F-test and AIC [Akaike 1974]. The model fits while using stelaCSF were significantly improved compared to fvvdpCSF ($F_{19,855} = 87.4889, p < .0005$). The AIC difference was 907.06. The prediction error of 3-4 dB is comparable to individual variations and the measurement errors found in the CSF studies (around 3 dB for nAFC). The detailed results, with per-dataset prediction errors and plots for each case, can be found on our project's web page1[0]. The results for the 5-dimensional CSFs are discussed in the next section.

6.2 Detailed results

In this section we discuss the predictions of stelaCSF and fvvdpCSF shown separately for each dataset. Both models were fitted to all datasets (models reported in the right-most column of Table 3). The predictions are plotted in Figures 8–18. We will use continuous lines for the predictions of stelaCSF and dashed lines for fvvdpCSF.

In Figure 8 we plot only the data points for Gabor patches from the ModelFest dataset [Watson and Ahumada 2005] and omit the stimuli with complex patterns as those cannot be predicted by a CSF alone. In Section 7.1 we show how such complex patterns can be predicted using stelaCSF combined with an energy model. Although

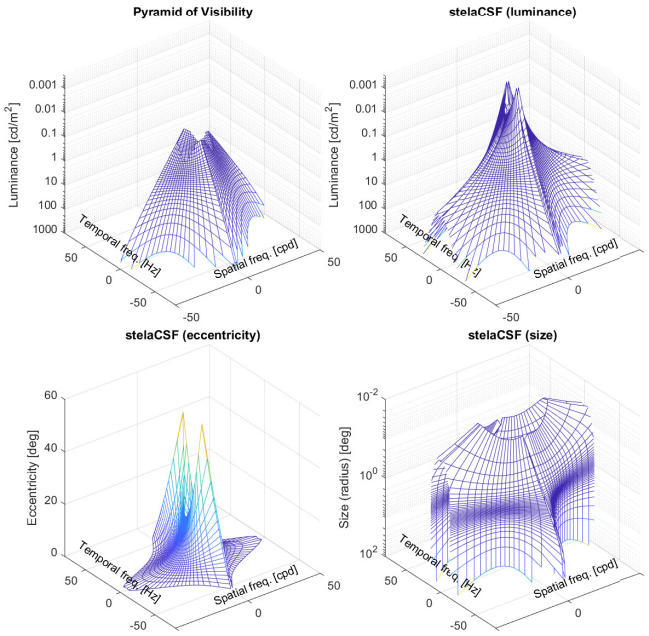

Fig. 7. The boundary of visibility for the pyramid of visibility [Watson and Ahumada 2016] (top, left) and stelaCSF (rest). The plots represent isosurfaces at which $S = 1$ (note that this is different from the pyramids shown in [Watson and Ahumada 2016]).

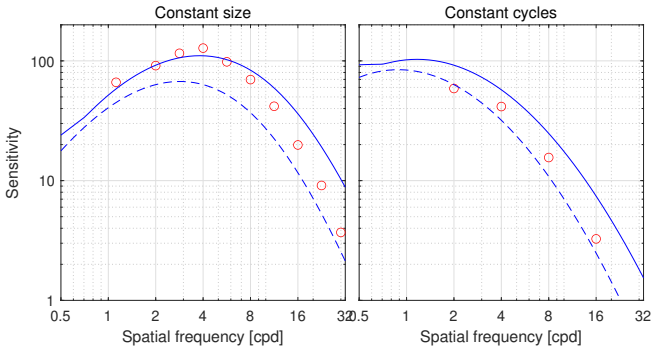

Fig. 8. Predictions for the ModelFest dataset [Watson and Ahumada 2005]. Continuous lines: stelaCSF; dashed lines: fvvdpCSF. This notation is followed in all figures.

ModelFest offers little challenge for most CSF models, it is one of the most robustly measured datasets, with measurements repeated across several laboratories. Therefore, we use this dataset to anchor the absolute predictions of the models (set $s_d = 1$ from Eq. (18) for this dataset).

HDR-VDP [Mantiuk et al. 2011] and HDR-CSF [Wuerger et al. 2020] datasets cover a wide range of luminance (from 0.002 to 10 000 cd/m^2) and stimulus sizes. The difference between the two datasets is that the HDR-VDP dataset was measured with a constant size stimulus (mostly 3 deg diameter) up to 200 cd/m^2, and the HDR-CSF dataset with a constant number of cycles (mostly 1 cycle) and

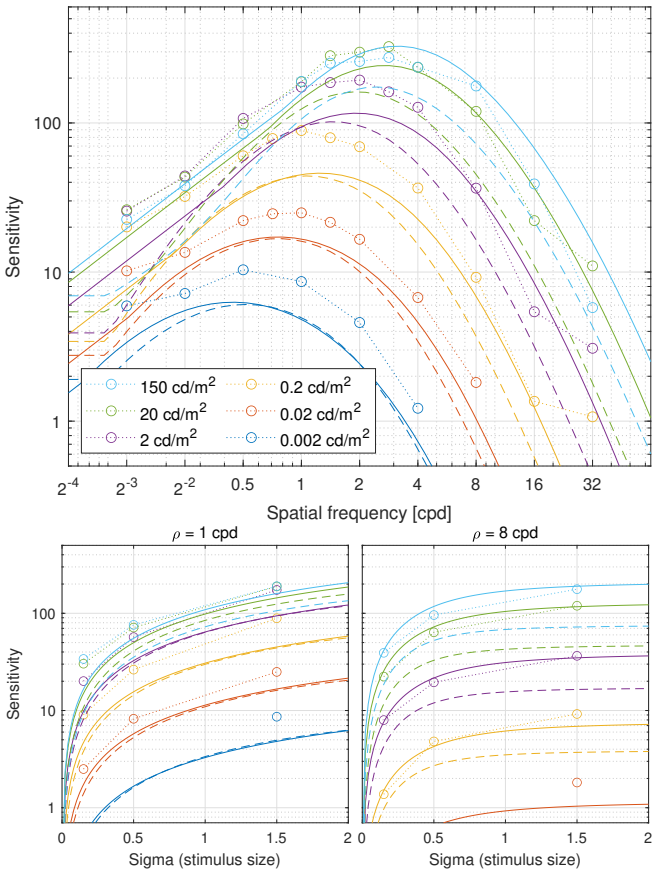

Fig. 9. Predictions for the HDR-VDP dataset [Mantiuk et al. 2011] for different frequencies (top) and stimulus sizes (bottom). The size is expressed as the radius of the Gaussian envelope.

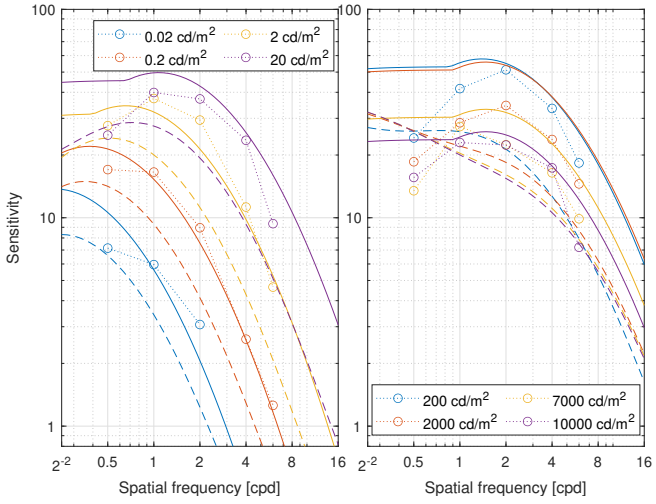

Fig. 10. Predictions for the HDR-CSF dataset [Wuerger et al. 2020]. Note that the contrast sensitivity drops after 200 cd/m^2 (plot on the right). The data was collected for constant-cycles Gabors.

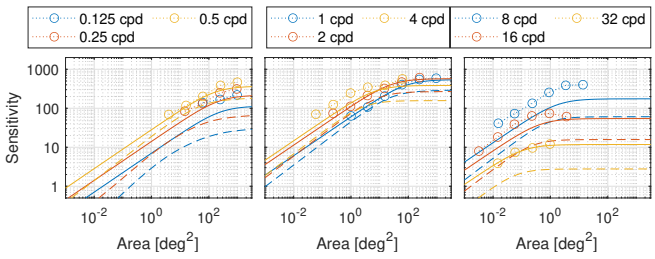

Fig. 11. Predictions for the Rovamo et al. dataset [1993].

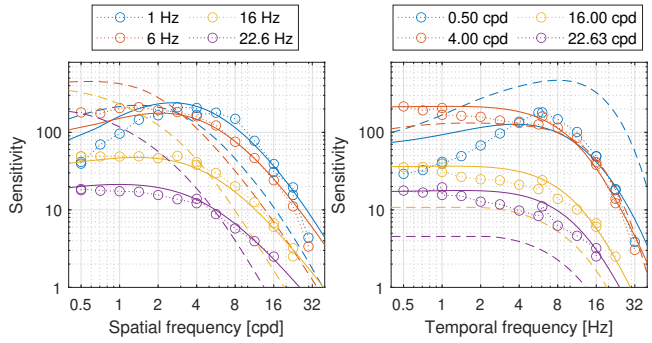

Fig. 12. Predictions for the Robson dataset [1966].

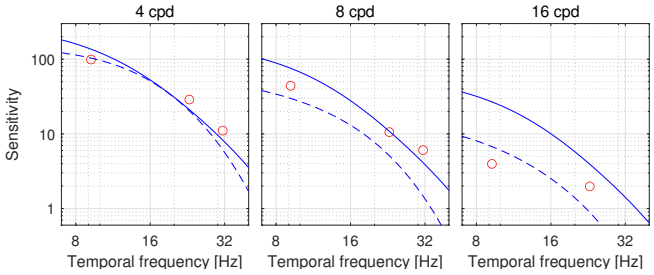

Fig. 13. Predictions for the Laird et al. dataset [2006].

up to 10 000 cd/m^2. Figure 9 and Figure 10 show that stelaCSF and fvvdpCSF well predict both datasets, including the sensitivity drop above 200 cd/m^2 (something that most CSF models fail to predict). fvvdpCSF predictions are less plausible for low frequencies at high luminance (Figure 10, right).

The Rovamo et al. dataset [Rovamo et al. 1993] demonstrates how sensitivity varies with the area of a stimulus. The predictions of both models, shown in Figure 11, do a good job of capturing the slope of the raise in sensitivity and the saturation beyond the critical area. The predictions for low and high frequencies are offset vertically from the data points, suggesting that the sensitivity changes with frequency cannot be predicted well for this dataset.

The datasets of Robson [1966] and Laird et al. [2006] contain variation of both spatial and temporal frequencies. The former was measured for a single observer but a large range of frequencies, while the latter was measured for 15 observers, but contains fewer data points. The predictions for the Robson dataset in Figure 12 show large discrepancies between the models, with stelaCSF predicting both band-pass and low-pass characteristic of spatio-temporal contrast sensitivity much better. The dataset of Laird et al. (Figure 13) was too limited to capture the differences between the models. We also attempted to predict the data from Kelly [1979a; 1979b], however, we found his data to be to different from all other datasets because of the stabilization of the stimuli on the retina used in his study.

The measurements of Snowden et al. [1995] extend that of Robson by capturing the changes in spatio-temporal contrast sensitivity with luminance. This dataset is important as it let us calibrate the parameters of the high-frequency ($\mathcal{T}$) channel of stelaCSF. The predictions, shown in Figure 14, indicate that both models can account for the changes of luminance, though stelaCSF predictions are much more accurate.

The remaining datasets capture the changes of sensitivity with eccentricity. Virsu and Rovamo measured the changes of sensitivity with eccentricity for static gratings in [1979] and then extended their measurement to 1 Hz and 18 Hz temporal modulations in [1982]. The predictions, shown in Figures 15 and 16, demonstrate that stelaCSF captures the trends well, though with lower accuracy at high spatial and temporal frequencies. Those prediction errors are consistent for all the datasets from Rovamo's lab. The very narrow band-pass characteristic shown in the left of Figure 15 could not be reproduced by our model as it is inconsistent with the other datasets. fvvdpCSF captured the changes of sensitivity at the eccentricity but with very large errors at low spatial frequencies.

The measurements of Wright and Johnston [1983] captured the changes of sensitivity with eccentricity for several combinations of both spatial and temporal frequencies, testing the 3 major dimensions of the models. The predictions, shown in Figure 17, indicate that a simple linear model used in stelaCSF (and in [Watson 2018]) provides an excellent fit to the data. A more complex cortical magnification model used in fvvdpCSF gave less plausible predictions.

Finally, in Figure 18, we show the predictions for the dataset of Anderson et al. [1991], who measured contrast sensitivity at four eccentricities but separately for different directions in the visual field (temporal, superior, nasal, inferior). They showed that the sensitivity drop with eccentricity is slower for the nasal direction, which motivated adding the position of the visual field to stelaCSF. The stelaCSF predictions align well with the data for low and medium frequencies, though they do not capture the rapid drop of sensitivity at high frequencies. We suspect that the prediction errors, which are larger than for the other datasets, are caused by the different criterion used to collect the data: discrimination of the motion direction instead of contrast detection (refer to Table 1). fvvdpCSF predicts a much shallower CSF and does not account for the position on the visual field.

6.3 Limitations

stelaCSF should be able to predict CFF measurements by finding the highest temporal frequency at which the sensitivity is equal to 1. However, we found the predictions to be not accurate enough to recommend using our model at high temporal frequencies. In Figure 19 we plot the predictions of stelaCSF and fvvdpCSF over the CFF data from [Krajancich et al. 2021]. While stelaCSF captures the

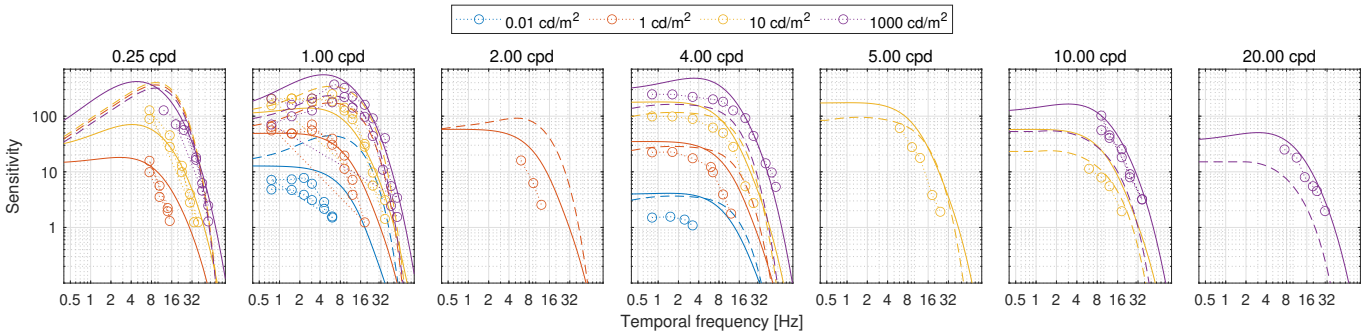

Fig. 14. Predictions for the Snowden et al. dataset [1995].

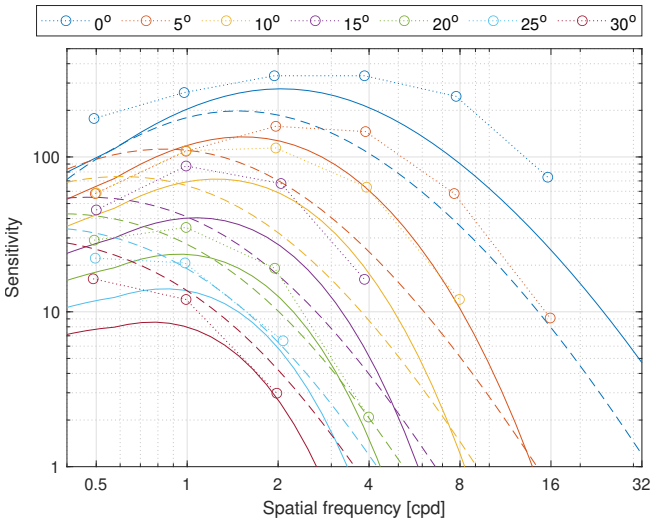

Fig. 15. Predictions for the Virsu and Rovamo dataset [1979].

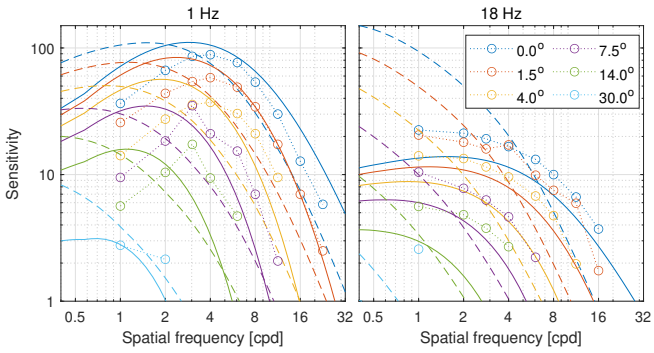

Fig. 16. Predictions for the Virsu et al. dataset [1982].

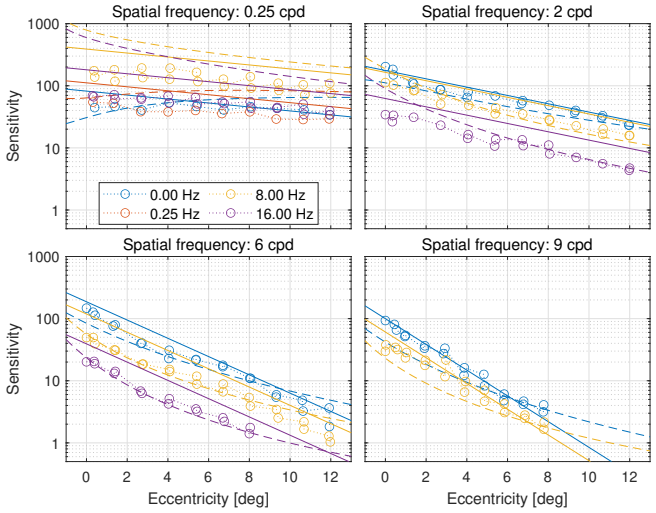

Fig. 17. Predictions for the Wright and Johnston dataset [1983].

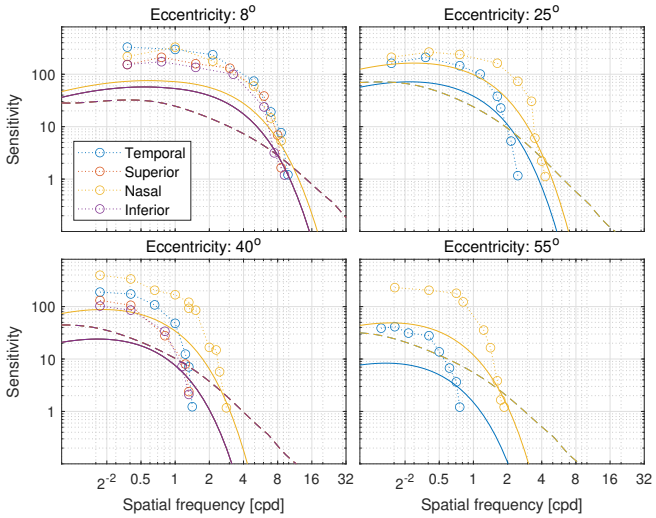

Fig. 18. Predictions for the Anderson et al. dataset [1991].

trends in the data, it does not predict the increase in CFF at medium eccentricities and it mispredicts the trend at very low frequencies (below 0.04 cpd). Such an increase in the CFF (and therefore in sensitivity) is not present in our contrast sensitivity datasets (see Figure 17) but was reported for other CFF measurements [Hartmann et al. 1979, Fig.3–4], [Tyler and Hamer 1990, Fig.7]. We suspect that

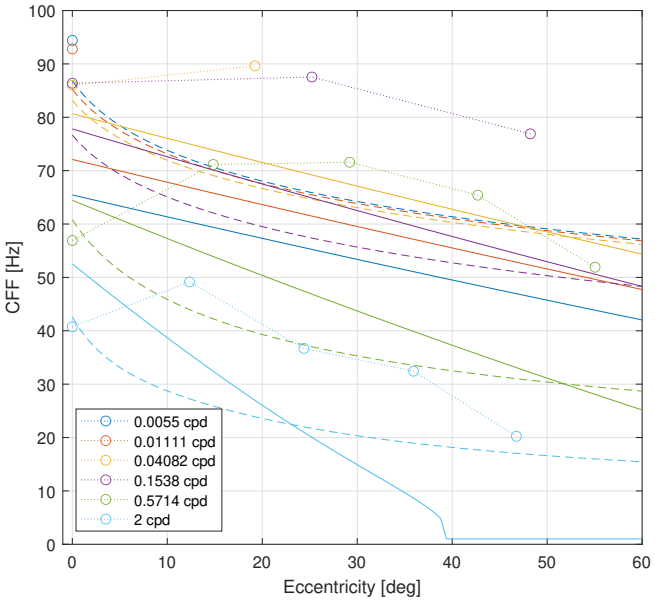

Fig. 19. Predictions for the Krajancich et al. [2021] dataset. Continuous lines: stelaCSF; dashed lines: fvvdpCSF. Neither of the models can predict the CFF data.

this characteristic can only be observed for low spatial and high temporal frequencies and higher luminance levels. The predictions of fvvdpCSF show that the cortical magnification component used in that model is also unable to predict this effect.

Koenderink et al. measured the CSF for moving patterns at eccentricities and at three levels or luminance [1978]. They found that while the sensitivity in the fovea decreased with luminance according to the de Vries-Rose behavior, it stayed constant at large eccentricities (up to 50 deg) according to the Weber law. However, this effect was not observed for static gratings in [Rovamo et al. 1995]. stelaCSF does not model interactions between luminance and eccentricity, and therefore cannot account for the effect observed by Koenderink et al.

As discussed in Section 6.1, existing CSF data does not allow for robust testing of CSF models using an independent subset of data not previously used for training. While this poses the risk of overfitting, this is mitigated by (a) fitting to all 11 datasets; and (b) using much fewer model parameters than data points. A CSF model should ideally be validated on a set of new measurements, randomly sampled from the 5D space and not used for training. However, collecting new measurements was outside the scope of this paper.

7 APPLICATIONS

In this section, we demonstrate three practical applications of our new CSF model to problems in computer graphics.

7.1 Contrast energy model

CSF models the detection thresholds of Gabor patches of isolated frequency and cannot directly predict thresholds for more complex patterns, which involve multiple frequencies. Due to this, we

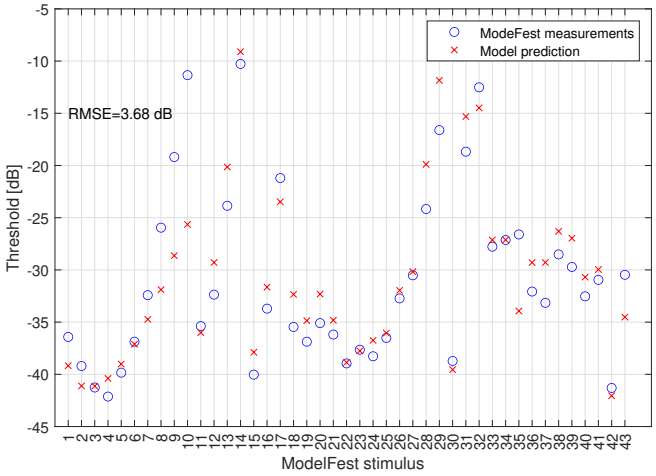

Fig. 20. Predictions of the contrast energy model for the complete set of 43 stimuli from the ModelFest dataset [Watson and Ahumada 2005].

used the ModelFest Gabor stimuli (1–14) when fitting CSF models, as other stimuli contain complex patterns which the CSF cannot predict alone. However, Watson and Ahumada showed that a simple contrast energy model could predict complex patterns near the detection threshold [2005]:

$$E(M,c) = \frac{1}{w^2} \sum_{(\rho_x,\rho_y)} \left(c\,\mathfrak{F}(M)(\rho_x,\rho_y)\,S(\rho,\omega_0,L_0,a,e)\right)^2 , \quad (20)$$

where M is the nominal modulation of a pattern (2D array with the values in the range -1 to 1), c is its contrast, and $\mathfrak{F}$ is the discrete Fourier transform. Summation is performed over all discrete horizontal and vertical coefficients ρ_x and ρ_y. w is the width of a pixel in visual degrees (1/120 deg for ModelFest). The orientation-invariant frequency ρ passed to stelaCSF (S from Eq. (17)) can be calculated as $\rho = \sqrt{\rho_x^2 + \rho_x^2}$. The spatial pooling is typically performed in the spatial, rather than frequency domain. However, as the exponent is equal to 2, we can use the Plancherel theorem to avoid the inverse Fourier transform and integrate the energy in the frequency domain. Because the contrast energy model performs spatial pooling, the area parameter a must be set to a value larger than the actual size of the stimulus. We set a to four times the size (8.53^2). The other parameters were set according to the ModelFest stimuli: $L_0 = 30$ cd/m^2, $\omega_0 = 0$ Hz, and $e = 0$ deg.

The model assumes that a pattern is detected when its contrast energy is above a certain threshold E_{thr}. Denoting the contrast detection threshold for i-th ModelFest stimulus as $c_{\text{det}}^{(i)}$, we have:

$$E_{\text{thr}} = E(M^{(i)}, c_{\text{det}}^{(i)}) = \left(c_{\text{det}}^{(i)}\right)^2 E(M^{(i)}, 1) \quad (21)$$

so that the threshold for each stimulus is:

$$c_{\text{det}}^{(i)} = \sqrt{\frac{E_{\text{thr}}}{E(M^{(i)}, 1)}} \quad (22)$$

E_{thr} can be found by optimizing for the smallest prediction error. There is also a closed-form solution to this optimization problem,

which can be computed as the squared geometric mean of the products $c_{\mathrm{GT}}^{(i)} \sqrt{E(M^{(i)}, 1)}$ across the dataset, where $c_{\mathrm{GT}}^{(i)}$ is the ground truth detection contrast.

The predictions of this simple model, shown in Figure 20, demonstrate that it can indeed predict ModelFest measurements for complex patterns (stimuli 15-43). The lower than expected detection thresholds at higher frequencies (stimuli 7–10), reflect the heightened sensitivity of stelaCSF at those frequencies (see Figure 8). The prediction error of 3.65 dB is higher than that of models fitted specifically to the ModelFest dataset. However, we only used the simplest form of the energy models from those proposed in [Watson and Ahumada 2005], and did not attempt to fit the CSF. Such a simple model can work well for low-contrast patterns, which are not affected by contrast masking. To model masking, the CSF should be integrated with more complex visual models, such as VDP [Daly 1992] or HDR-VDP [Mantiuk et al. 2011].

7.2 Flicker Detection

In this Section, we demonstrate a practical use of stelaCSF as a component of a flicker detection algorithm. Flicker may arise in display presentation under several conditions, but is chiefly due to large luminance changes at low refresh rates - both factors that affect contrast sensitivity. The threshold visibility of flicker is often modeled using the *critical flicker fusion* frequency (CFF, see Section 3).

Denes and Mantiuk [2020] propose an algorithm to detect perceptible flicker in images. One component of obtaining the desired prediction is the decomposition of flickering image pairs in the spatial frequency domain, followed by filtering through a CSF. Denes and Mantiuk use the Pyramid of Visibility (PoV) [Watson and Ahumada 2016] in their implementation. Notably, this model is designed to account only for high frequencies (>10 Hz or >10 cpd) and, in its original form, does not model eccentricity. To address these limitations, we modified Denes and Mantiuk's flicker detector to use stelaCSF. Apart from changing the CSF model, we modified the spatial frequency calculation for the Laplacian pyramid to the updated formula from [Mantiuk et al. 2021]. Finally, we reduce the three fitted parameters of the PoV used in their work to a single parameter controlling the absolute sensitivity of the model.

In order to test the modified flicker predictor, we compare against results presented in the validation experiment of the recent work of Krajancich et al. [2021]. This study gathered CFF data for a variety of spatial and temporal frequencies, luminance levels, and eccentricities. The authors also present a flicker detection model, which is obtained by mathematical modeling of the parameters used to generate the flickering Gabor stimuli used in their study. The model of Krajancich et al. accounts for the background luminance, but ignores image content. Despite this limitation, in their validation study the authors show improvement over traditional flicker-unaware metrics like SSIM, PSNR, and VMAF when predicting the visibility of flicker introduced by adding a Gabor over a static image. Alternative flicker-aware metrics, such as that proposed by Denes and Mantiuk, were not considered for comparison as they did not model eccentricity.

As shown in Figure 21, with the introduction of stelaCSF, the algorithm is able to accurately predict the detection of flicker for the conditions studied in the experiment, including accounting for the previously missing effect of eccentricity.

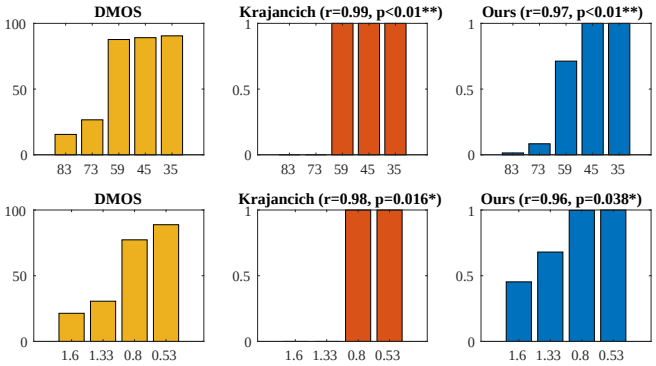

Fig. 21. A flicker detection algorithm [Denes and Mantiuk 2020] was extended to use stelaCSF, and is used to predict the validation results of work by Krajancich et al. [2021]. The top row shows condition I, *unseen stimulus*, where a fixed Gabor is presented flickering at various frequencies. The bottom row shows condition II, f_s *dependence*, where flickering Gabors of fixed size but varying spatial frequency are shown at a single temporal frequency. We reproduce the mean DMOS score of the user study on the left, the result obtained by the authors' model in the center, and show the novel results of our modified algorithm on the right.

Note that although we obtain performance similar to the Krajancich et al. model as measured by Pearson's r and p values, the flicker detector using stelaCSF is much more general. The new method can predict flicker for any image and distortion type, and is not limited to Gabors. Furthermore, unlike the model presented by Krajancich et al., this flicker detector is not limited to binary outputs (visible or not), and can disambiguate between affected areas of an image and give localized probability estimates, as illustrated in Figure 22.

Fig. 22. A prediction of our CFF detector based on the work of Denes and Mantiuk. Original scene from Sitzmann et al..

7.3 Foveated rendering

As display parameters such as resolution, luminance and frame rate continue to increase, computational resources are strained to produce high-quality results at higher sampling rates. Foveated rendering aims to decrease the cost to render or transmit information by reducing the detail level at which it is generated in areas away from the users' gaze. Such methods are typically based on models of peripheral vision [Guenter et al. 2012; Patney et al. 2016].

While some modern methods like the one presented by Tursun et al. [2019] employ explicit CSF-aware optimization processes to generate optimal simplification profiles, these can be computationally expensive. Instead, most techniques employ a fixed fall-off profile to reduce the sampling rate with eccentricity, based on the results of

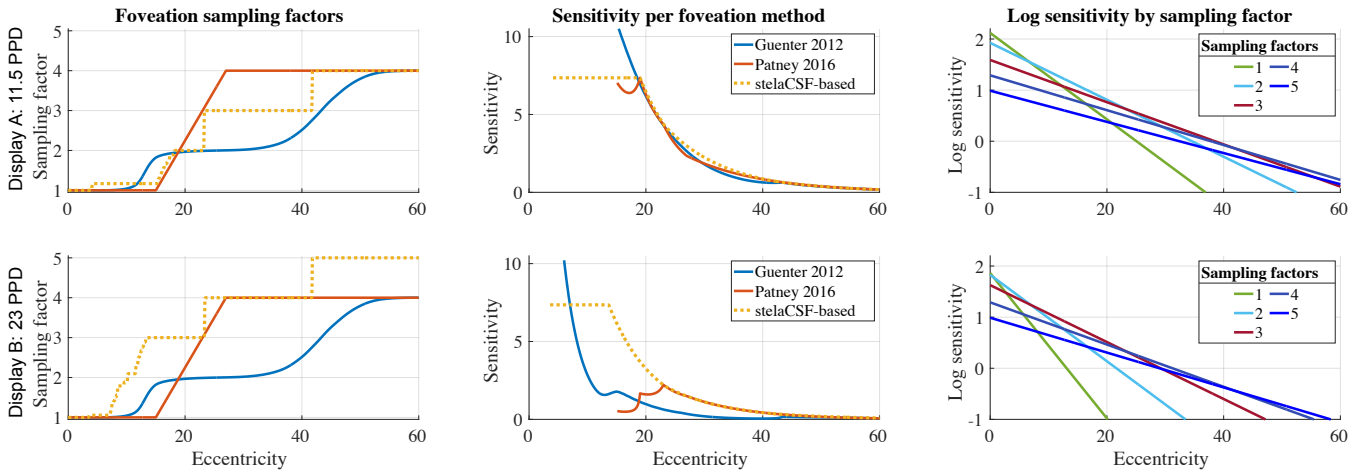

Fig. 23. In the top row, we show the foveation profiles presented in prior art for a display with 11.5 ppd, contrasted against our automatically generated profile (top-left). Note that at low eccentricities, foveation is not applied. Next, the sensitivity values for the foveated regions of each method is calculated (top-middle). The sensitivity predicted by stelaCSF for each sampling factor is shown (top-right). The bottom row repeats this simulation for a display with 23 ppd, demonstrating that previous foveation methods become suboptimal with a change in resolution.

a perceptual experiment. Although this method can produce good results, our work shows that significant changes in peripheral sensitivity can occur due to changes in ancillary variables like display resolution. These changes could degrade the performance of such manually-developed foveation profiles.

To estimate the perceptual impact of foveated rendering, we first obtained the foveation profiles presented by Patney et al. [2016] and Guenter et al. [2012] as presented in the former's publication (Figure 23 top left). We then model the frequency profile of the hardware used in their work (Oculus Dev Kit 2, 110° FOV, 11.5 ppd, estimated 100 cd/m^2 luminance). We calculate the Nyquist frequency for each sampling factor as $\hat{\rho}_i = \hat{\rho}_{i-1}/2$ and $\hat{\rho}_1 = \text{ppd}/2$, obtaining values of 5.75, 2.9, 1.4, and 0.7 cpd. We use stelaCSF to estimate the sensitivity to the Nyquist frequency of each sampling level (Figure 23 top right), and calculate the sensitivity of blended layers through linear interpolation as follows: $S^{\alpha}_{i/i+1} = (\alpha/S_i + (1-\alpha)/S_{i+1})^{-1}$, since we are blending contrast, which is the inverse of sensitivity. This allows us to calculate the sensitivity profile of both foveated rendering methods at each eccentricity value (Figure 23 top middle).

As an alternative, we use stelaCSF to automatically generate a foveation profile by finding the highest simplification magnitude at each eccentricity that stays below a given threshold of sensitivity. We chose the threshold as the maximum value calculated for Patney and colleagues' [2016] method, 7.3. Although their original method was limited to 4 sampling steps, we are able to simulate more aggressive downsampling, choosing to go up to 5 steps in this simulation. First, we note that the foveation profiles we examine are similar to our automatic curve when examined at 11.5 ppd. However, this changes when a different display is simulated using our framework: in the bottom row of Figure 23, we repeat the same steps for a display with double the ppd (23) of that used by the authors. This increases the Nyquist frequency, reducing the sensitivity across the periphery, making the previous foveation profiles overly conservative. Our stelaCSF-based foveation is able to automatically generate a perceptually optimal curve, predicted to be spatially indistinguishable from the original, and enabling computational gains when considered in tandem with practical engineering limitations.

8 CONCLUSIONS

The CSF is a practical model of the visual system that can help in assessing the visibility of artifacts. It is also a key component of more complex visual models, or visual metrics. However, by itself contrast sensitivity is a reliable measure only for patterns that are well specified by its main characteristics: spatial and temporal frequency, luminance, size and eccentricity. In this paper we provide a practical model that accounts for all these dimensions and explains a large volume of psychophysical data.

While it is possible to fit any function to the data, an arbitrary function is unlikely to explain the sparsely measured space of contrast sensitivity. For that reason, stelaCSF builds on existing findings and visual models, such as Rovamo et al.'s relation between sensitivity and the product of square-root area and frequency, or Watson's observation that the effect of eccentricity can be explained by linear relations. Our selection of models and relevant relations is based on both empirical evidence reflected in 11 examined datasets, and findings reported in the vision literature.

We believe our model could be improved in terms of prediction at higher temporal frequencies, such as those collected in CFF measurements. Our model also does not consider chromatic contrast sensitivity [Mantiuk et al. 2020; Wuerger et al. 2020].

stelaCSF opens possibilities for numerous applications, especially in the space of novel VR/AR technologies. These often suffer from spatio-temporal artifacts, which can be characterized and then assessed in the 5-dimensional space of stelaCSF. We demonstrate this on two example applications in detecting visible flicker and estimating sampling factors for foveated rendering.

ACKNOWLEDGMENTS

We would like to thank anonymous reviewers, Özgür Yöntem, Minjung Kim and Romain Bachy for their valuable comments. This project has received funding from the European Research Council (ERC) under the European Union's Horizon 2020 research and innovation programme (grant agreement N° 725253–EyeCode).

REFERENCES

Albert Ahumada and Heidi A. Peterson. 1992. Luminance-model-based DCT quantization for color image compression. In *SPIE 1666, Human vision, visual processing, and digital display III.*

Albert Ahumada, Jihyun Yeonan-Kim, and Andrew B. Watson. 2018. A Dual Channel Spatial-Temporal Detection Model. *Electronic Imaging* 2018, 14 (2018), 1–4.

Hirotugu Akaike. 1974. A new look at the statistical model identification. *IEEE transactions on automatic control* 19, 6 (1974), 716–723.

S. J. Anderson, K. T. Mullen, and R. F. Hess. 1991. Human peripheral spatial resolution for achromatic and chromatic stimuli: limits imposed by optical and retinal factors. *The Journal of Physiology* 442 (1991), 47–64.

Pontus Andersson, Jim Nilsson, Tomas Akenine-Möller, Magnus Oskarsson, Kalle Åström, and Mark D Fairchild. 2020. FLIP: A Difference Evaluator for Alternating Images. *Proc. ACM Comput. Graph. Interact. Tech.* 3, 2 (2020), 15–1.

Peter G. J. Barten. 1999. *Contrast sensitivity of the human eye and its effects on image quality.* SPIE Press.

Peter G. J. Barten. 2003. Formula for the contrast sensitivity of the human eye. In *Proc. SPIE 5294, Image Quality and System Performance.* 231–238.

Christina A. Burbeck and D. H. Kelly. 1980. Spatiotemporal characteristics of visual mechanisms: excitatory-inhibitory model. *Journal of the Optical Society of America* 70, 9 (September 1980), 1121–1126.

Fergus W Campbell and John G Robson. 1968. Application of Fourier analysis to the visibility of gratings. *The Journal of physiology* 197, 3 (1968), 551–566.

Scott J Daly. 1992. Visible differences predictor: an algorithm for the assessment of image fidelity. In *SPIE 1666, Human Vision, Visual Processing, and Digital Display III.* International Society for Optics and Photonics, 2–15.

Scott J. Daly. 1998. Engineering observations from spatiovelocity and spatiotemporal visual models. In *SPIE 3299, Human Vision and Electronic Imaging.* 180–191.

H De Lange. 1952. Experiments on flicker and some calculations on an electrical analogue of the foveal systems. *Physica* 18, 11 (1952), 935–950.

Gyorgy Denes and Rafał K. Mantiuk. 2020. Predicting visible flicker in temporally changing images. *Electronic Imaging* 2020, 11 (2020), 233–1.

M. A. Georgeson and G. D. Sullivan. 1975. Contrast constancy: deblurring in human vision by spatial frequency channels. *The J. of Physiology* 252, 3 (1975), 627–656.

Brian Guenter, Mark Finch, Steven Drucker, Desney Tan, and John Snyder. 2012. Foveated 3D graphics. *ACM Transactions on Graphics (TOG)* 31, 6 (2012), 1–10.

S.T. Hammett and A.T. Smith. 1992. Two temporal channels or three? A re-evaluation. *Vision Research* 32, 2 (February 1992), 285–291.

E. Hartmann, B. Lachenmayr, and H. Brettel. 1979. The peripheral critical flicker frequency. *Vision Research* 19, 9 (jan 1979), 1019–1023. https://doi.org/10.1016/0042-6989(79)90227-X

D. H. Kelly. 1979a. Motion and vision. I. Stabilized images of stationary gratings. *Journal of the Optical Society of America* 69, 9 (1979), 1266–1274.

D. H. Kelly. 1979b. Motion and vision. II. Stabilized spatio-temporal threshold surface. *Journal of Optical Society of America* 69, 10 (1979), 1340–1349.

Jan J. Koenderink, Maarten A. Bouman, Albert E. Bueno de Mesquita, and Sybe Slappendel. 1978. Perimetry of contrast detection thresholds of moving spatial sine wave patterns IV The influence of the mean retinal illuminance. *Journal of the Optical Society of America* 68, 6 (June 1978), 860 – 865.

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 Trans. Graph.* 40, 4, Article 47 (July 2021), 11 pages.

Justin Laird, Mitchell Rosen, Jeff Pelz, Ethan Montag, and Scott Daly. 2006. Spatio-velocity CSF as a function of retinal velocity using unstabilized stimuli. In *SPIE 6057, Human Vision and Electronic Imaging XI.*

G. E. Legge and J. M. Foley. 1980. Contrast masking in human vision. *Journal of the Optical Society of America* 70, 12 (December 1980), 1458–1471.

Rafal Mantiuk, Scott J Daly, Karol Myszkowski, and Hans-Peter Seidel. 2005. Predicting visible differences in high dynamic range images: model and its calibration. In *Human Vision and Electronic Imaging X*, Vol. 5666. International Society for Optics and Photonics, 204–214.

Rafał K. Mantiuk, S. Daly, and L. Kerofsky. 2008. Display adaptive tone mapping. *ACM Transactions on Graphics* 27, 3 (2008), 1–10.

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 Transaction on Graphics* 40, 4 (2021), 1–19.

Rafał K. Mantiuk, Kil Joong Kim, Allan G. Rempel, and Wolfgang Heidrich. 2011. HDR-VDP-2: A calibrated visual metric for visibility and quality predictions in all luminance conditions. *ACM Transactions on Graphics* 30, 4 (July 2011), 1–14.

Rafał K. Mantiuk, Minjung Kim, Maliha Ashraf, Qiang Xu, M. Ronnier Luo, Jasna Martinovic, and Sophie Wuerger. 2020. Practical Color Contrast Sensitivity Functions for Luminance Levels up to 10000 cd/m^2. In *Color and Imaging Conference*, Vol. 2020. Society for Imaging Science and Technology, 1–6.

W. H. Merigan, L. M. Katz, and J. H. Maunsell. 1991. The effects of parvocellular lateral geniculate lesions on the acuity and contrast sensitivity of macaque monkeys. *The Journal of Neuroscience* 11, 4 (April 1991), 994–1001.

Juvi Mustonen, Jyrki Rovamo, and Risto Näsänen. 1993. The effects of grating area and spatial frequency on contrast sensitivity as a function of light level. *Vision Research* 33, 15 (October 1993), 2065–2072.

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 (2016), 1–12.

D. G. Pelli, Melanie Palomares, and Najib J. Majaj. 2004. Crowding is unlike ordinary masking: Distinguishing feature integration from detection. *Journal of Vision* 4, 12 (dec 2004), 12. https://doi.org/10.1167/4.12.12

John G. Robson. 1966. Spatial and temporal contrast-sensitivity functions of the visual system. *Journal of Optical Society of America* 56, 8 (1966), 1141–1142.

Ankit Rohatgi. 2021. Webplotdigitizer: Version 4.5. https://automeris.io/WebPlotDigitizer

Jyrki Rovamo, Olavi Luntinen, and Risto Näsänen. 1993. Modelling the dependence of contrast sensitivity on grating area and spatial frequency. *Vision Research* 33, 18 (December 1993), 2773–2788.

Jyrki Rovamo, Juvi Mustonen, and Risto Näsänen. 1995. Neural modulation transfer function of the human visual system at various eccentricities. *Vision Research* 35, 6 (March 1995), 767–774.

Robert L. Savoy and John J. McCann. 1975. Visibility of low-spatial-frequency sine-wave targets: Dependence on number of cycles. *Journal of the Optical Society of America* 65, 3 (March 1975), 343–350.

Otto H. Schade. 1956. Optical and photoelectric analog of the eye. *Journal of Optical Society of America* 46, 9 (1956), 721–739.

Vincent Sitzmann, Ana Serrano, Amy Pavel, Maneesh Agrawala, Diego Gutierrez, Belen Masia, and Gordon Wetzstein. 2018. Saliency in VR: How do people explore virtual environments? *IEEE transactions on Visualization and Computer Graphics* (2018).

Robert J. Snowden, Robert F. Hess, and Sarah J. Waugh. 1995. The Processing of Temporal Modulation at Different Levels of Retinal Illuminance. *Vision Research* 35, 6 (1995), 775–789.

Hans Strasburger, Ingo Rentschler, and Martin Jüttner. 2011. Peripheral vision and pattern recognition: A review. *Journal of Vision* 11, 5 (2011), 1–82.

Okan Tarhan Tursun, Elena Arabadzhiyska-Koleva, Marek Wernikowski, Radosław Mantiuk, Hans-Peter Seidel, Karol Myszkowski, and Piotr Didyk. 2019. Luminance-contrast-aware foveated rendering. *ACM Transactions on Graphics* 38 (2019).

Christopher W. Tyler and Russell D. Hamer. 1990. Analysis of visual modulation sensitivity. IV. Validity of the Ferry–Porter law. *Journal of Optical Society of America A* 7, 4 (1990), 743–758.

Christopher W Tyler and Russell D Hamer. 1993. Eccentricity and the Ferry–Porter law. *Journal of Optical Society of America A* 10, 9 (1993), 2084–2087.

Peter Vangorp, Karol Myszkowski, Erich W. Graf, and Rafał K. Mantiuk. 2015. A model of local adaptation. *ACM Transactions on Graphics* 34, 6 (oct 2015), 1–13. https://doi.org/10.1145/2816795.2818086

V. Virsu and J. Rovamo. 1979. Visual resolution, contrast sensitivity, and the cortical magnification factor. *Experimental Brain Research* 37, 3 (1979), 475–494.

Veijo Virsu, Jyrki Rovamo, and Pentti Laurinen. 1982. Temporal Contrast Sensitivity and Cortical Magnification. *Vision Research* 22 (1982), 1211–1217.

Andrew B. Watson. 2000. Visual detection of spatial contrast patterns: Evaluation of five simple models. *Optics Express* 6, 1 (2000), 12–33.

Andrew B. Watson. 2018. The field of view, the field of resolution, and the field of contrast sensitivity. *Journal of Perceptual Imaging* 1, 1 (January 2018), 1–11.

Andrew B. Watson and Albert J. Ahumada. 2005. A standard model for foveal detection of spatial contrast. *Journal of Vision* 5, 9 (2005), 717–740.

Andrew B. Watson and Albert J. Ahumada. 2016. The pyramid of visibility. *Electronic Imaging* 16 (2016), 37–42.

M. J. Wright and A. Johnston. 1983. Spatiotemporal contrast sensitivity and visual field locus. *Vision Research* 23, 10 (1983), 983–989.

Sophie Wuerger, Maliha Ashraf, Minjung Kim, Jasna Martinovic, María Pérez-Ortiz, and Rafał K. Mantiuk. 2020. Spatio-chromatic contrast sensitivity under mesopic and photopic light levels. *Journal of Vision* 20, 4 (April 2020), 23.

Shinyoung Yi, Daniel S. Jeon, Ana Serrano, Se-Yoon Jeong, Hui-Yong Kim, Diego Gutierrez, and Min H. Kim. 2022. Modelling Surround-aware Contrast Sensitivity for HDR Displays. *Computer Graphics Forum* 41, 1 (feb 2022), 350–363.

Wenjun Zeng, Scott Daly, and Shawmin Lei. 2002. An overview of the visual optimization tools in JPEG 2000. *Signal Processing: Image Communication* 17 (2002).