

Spatial summation of broadband contrast

Bruno Richard

Department of Mathematics and Computer Science,
Rutgers University, Newark, NJ, USA



Bruce C. Hansen

Department of Psychological and Brain Sciences,
Neuroscience Program, Colgate University,
Hamilton, NY, USA

Aaron P. Johnson

Department of Psychology, Concordia University,
Montreal, Quebec, Canada

Patrick Shafto

Department of Mathematics and Computer Science,
Rutgers University, Newark, NJ, USA

Spatial summation of luminance contrast signals has historically been psychophysically measured with stimuli isolated in spatial frequency (i.e., narrowband). Here, we revisit the study of spatial summation with noise patterns that contain the naturalistic $1/f^\alpha$ distribution of contrast across spatial frequency. We measured amplitude spectrum slope (α) discrimination thresholds and verified if sensitivity to α improved according to stimulus size. Discrimination thresholds did decrease with an increase in stimulus size. These data were modeled with a summation model originally designed for narrowband stimuli (i.e., single detecting channel; Baker & Meese, 2011; Meese & Baker, 2011) that we modified to include summation across multiple—differently tuned—spatial frequency channels. To fit our data, contrast gain control weights had to be inversely related to spatial frequency ($1/f$); thus low spatial frequencies received significantly more divisive inhibition than higher spatial frequencies, which is a similar finding to previous models of broadband contrast perception (Haun & Essock, 2010; Haun & Peli, 2013). We found summation across spatial frequency channels to occur prior to summation across space, channel summation was near linear and summation across space was nonlinear. Our analysis demonstrates that classical psychophysical models can be adapted to computationally define visual mechanisms under broadband visual input, with the adapted models offering novel insight on the integration of signals across channels and space.

an increase in its area, and occurs for stimuli presented at and above contrast threshold levels (Baker & Meese, 2011; Campbell & Green, 1965; Graham, 1977; Graham & Robson, 1987; Graham, Robson, & Nachmias, 1978; Graham & Sutter, 1998; Kersten, 1984; Landy & Oruç, 2002; Legge, 1984; Meese, 2004; Meese & Baker, 2011; Meese, Hess, & Williams, 2005; Meese & Summers, 2007; Robson & Graham, 1981; Summers, Baker, & Meese, 2015). Computationally, spatial summation is described as a multistage process that begins with spatial filtering (i.e., a filter narrowly tuned for spatial frequency and orientation), followed by nonlinear transduction, linear summation and probably summation, where each stage operates over a progressively larger area of the retina (Baker & Meese, 2011; Foley, Varadharajan, Koh, & Farias, 2007; Meese, 2004; Meese & Baker, 2011; Meese & Summers, 2007; Wilson & Gelb, 1984). This structure of summation is an excellent foundation that accounts for psychophysical and neurophysiological summation effects across eye and space for phase congruent and incongruent stimuli—excluding modality specific terms like interocular suppression for binocular combination and phase selective channels (Baker & Meese, 2011; Cunningham, Baker, & Peirce, 2017; Georgeson, Wallis, Meese, & Baker, 2016; Meese, 2004; Meese & Baker, 2011; Meese, Georgeson, & Baker, 2006; Richard, Chadnova, & Baker, 2018). Summation models (including recent implementations; Baker & Meese, 2011, Meese & Baker, 2011) have been developed explicitly with narrowband stimuli (i.e., sinusoidal gratings), and thus can only describe the response of a single detecting channel to a stimulus. Yet, the retinal image formed by real-world environ-

Introduction

Spatial summation of luminance contrast signals describes an increase in sensitivity to a stimulus given

Citation: Richard, B., Hansen, B. C., Johnson, A. P., & Shafto, P. (2019). Spatial summation of broadband contrast. *Journal of Vision*, 19(5):16, 1–19, <https://doi.org/10.1167/19.5.16>.

<https://doi.org/10.1167/19.5.16>

Received September 17, 2018; published May 17, 2019

ISSN 1534-7362 Copyright 2019 The Authors



ments is broadband: it contains contrast across a broad range of spatial frequencies and orientations. This means that multiple channels with different tuning are simultaneously active, and their outputs weighted by the interdependent responses of similarly and dissimilarly tuned channels (Cass, Stuit, Bex, & Alais, 2009; Schwartz & Simoncelli, 2001). To understand how these channels operate (e.g., spatially sum) in naturalistic environments, it is important to measure psychophysical effects with stimuli that better represent the typical input received by the visual system (i.e., broadband images). This, in turn, can guide how psychophysical models of vision may be adjusted to describe how vision operates in the real world (Bex, Mareschal, & Dakin, 2007; Hansen et al., 2015; Hansen & Hess, 2012; Haun & Peli, 2013; Legge & Foley, 1980; Meese & Holmes, 2010; Petrov, Carandini, & McKee, 2005; Schwartz & Simoncelli, 2001).

Increasing the complexity of a stimulus from a narrowband image, which contains a single spatial frequency and orientation, to images that contain more than one spatial frequency or orientation (i.e., broadband images) has a measurable impact on perception. This has been demonstrated psychophysically with studies on cross-orientation suppression (Meese & Holmes, 2007, 2010; Roeber, Wong, & Freeman, 2008), the horizontal effect (Essock, Haun, & Kim, 2009; Hansen, Essock, Zheng, & DeFord, 2003; Hansen et al., 2015), broadband masking (Hansen & Hess, 2012), amplitude spectrum slope discrimination (Hansen & Hess, 2006; Johnson, Richard, Hansen, & Elleberg, 2011; Knill, Field, & Kersten, 1990), and perceived contrast in natural images (Haun & Peli, 2013). Importantly, those studies have repeatedly demonstrated that stimuli containing contrast across a broad range of spatial frequencies and orientations can produce interactive processes that alter observer sensitivity to spatial frequency and/or orientation. The few that explored broadband contrast perception computationally demonstrated that a classical contrast gain control transducer with an additional divisive term for the activation of differently tuned channels is sufficient to capture observer responses to naturalistic stimuli. Indeed, this computational approach was successfully implemented to describe observer responses for traditional psychophysical tasks of visual masking (Alam, Vilankar, Field, & Chandler, 2014; Hansen et al., 2015), perceived contrast (Haun & Peli, 2013), and binocular summation (Huang & Dai, 2018). These studies, which implement traditional psychophysical paradigms with naturalistic stimuli, have made important contributions to the development of a model of broadband contrast perception. However, there remain fundamental features of visual processing to explore with broadband stimuli. Here, we investigate how the responses of differently tuned spatial frequency

channels may sum together over space, and whether their summation leads to any improvement in observer sensitivity. That is, we investigate whether sensitivity to broadband stimuli is subject to spatial summation effects.

An increase in area of a broadband stimulus alters a stimulus in a manner that differs from an increase in area of a sinusoidal grating. While increasing the area of a sinusoidal grating will only increase the number of cycles displayed, making the stimulus increasingly narrowband, an increase in size of a broadband stimulus will add low spatial frequencies to the image.¹ For natural scenes, contrast is unevenly represented across spatial frequency: the Fourier amplitude spectrum of natural images falls inversely with spatial frequency, defined as

$$\text{amplitude} \propto \frac{1}{f^\alpha} \quad (1)$$

where f is spatial frequency and the exponent (α) defines the rate of descent in amplitude. On average, the value of α approximates 1.0, but ranges between 0.6 and 1.6 across wide sets of natural scenes (Billock, 2000; Burton & Moorhead, 1987; Hansen & Essock, 2005; Tolhurst, Tadmor, & Chao, 1992; van der Schaaf & van Hateren, 1996). This means that natural images possess more contrast at low spatial frequencies than high, with the relative difference in contrast indexed by α . As images with steeper α have a larger representation of low spatial frequency contrast, an increase in stimulus size (thus adding low spatial frequencies to the image), may exert a larger influence on discrimination thresholds than images with shallower α . It is unclear how this additional low spatial frequency content may influence observer perception, as masking studies that use naturalistic masks have demonstrated that percepts based on low spatial frequencies are disproportionately suppressed compared to high spatial frequencies (Bex, Solomon, & Dakin, 2009; Haun & Essock, 2010; Webster & Miyahara, 1997). The added low spatial frequency content from an increase in stimulus size may therefore have little influence on sensitivity to broadband stimuli. As it is unclear how the α of an image may modulate spatial summation (if at all), broadband spatial summation should be measured with images that range in α in order to adequately capture the summation process potentially completed by the visual system under naturalistic scenarios.

There is, additionally, a question in regard to how to best measure spatial summation with broadband stimuli. Spatial summation is typically defined as a decrease in contrast detection or discrimination thresholds, measurements that are challenging to make with broadband stimuli. We can, however, measure sensitivity to the distribution of broadband contrast across spatial frequency with an amplitude spectrum

slope (α) discrimination task (Field, 1987; Hansen & Hess, 2006; Johnson et al., 2011; Tolhurst & Tadmor, 1997). In this task, observers are asked to discriminate between images (e.g., noise or natural scenes) that differ in α , which can serve as a proxy of broadband contrast sensitivity. Discriminability is known to depend on the reference α : discrimination thresholds are generally lowest when the reference stimulus has an α near the typical values of natural scene images (e.g., between 1.0–1.3; Ellefberg, Hansen, & Johnson, 2012; Field, 1987; Hansen & Hess, 2006; Johnson et al., 2011; Knill et al., 1990). These studies were conducted with small (1° – 2°) stimuli presented at fovea or parafovea. Whether tuning to α persists for larger stimulus sizes is uncertain as the only other study that measured α discrimination thresholds with large stimuli ($\sim 10^\circ$) found no indication the typical peak in sensitivity for α s near 1.0–1.3 (Thomson & Foster, 1997). Amplitude spectrum slope discrimination thresholds were measured with images of natural scenes and in their dataset, sensitivity to α was actually worse when the reference image α was closest to its original value, which indicates that tuning to α may be altered by the size of the stimulus.

Motivated by the above, we first verify whether sensitivity to broadband stimuli that vary in α improves as a function of stimulus size. Second, we use these data to build a model of broadband spatial summation in a similar fashion to previous computational descriptions of broadband contrast perception. We measure α discrimination thresholds for five different reference α values to stimuli of increasing size. We find that discrimination thresholds decreased monotonically as a function of stimulus area, and interestingly, this decrease was not modulated by the reference α . This means that the increase in stimulus size did not alter tuning to α : thresholds remained lowest for α values between 1.0–1.3. To explain these findings, we develop a computational framework that adapts a narrowband spatial summation model (Baker & Meese, 2011; Foley et al., 2007; Meese & Baker, 2011; Meese & Summers, 2007) to generate responses for broadband and naturalistic stimuli. Specifically, we explore how modifications to the summation of the responses over multiple spatial frequency channels and the linearity of summation affect the ability of a spatial summation model to fit the broadband spatial summation results obtained here.

Methods

Participants

Seven volunteers (five women, two men) between the ages of 19 and 23 (median age = 21 years) participated

in the experiment. All were experienced psychophysics observers with normal or corrected-to-normal visual acuity. Informed consent was obtained from all participants, and all were treated in accordance the Tri-Council Policy Statement: Ethical Conduct for Research Involving Humans, and approved by Concordia University Human Research Ethics Committee (Certificate: 10000119).

Apparatus

Stimuli were presented on a 22.5 in. ViewSonic (G225fB) monitor driven by an Apple Mac Pro (2×2.66 GHz processor) equipped with 8 GB of RAM and a 1 GB PCIe $\times 16$ ATI Radeon HD 5770 Graphics card with 10-bit grayscale resolution.² Stimuli were displayed using a linearized look-up table, generated by calibrating with a Color-Vision Spyder3 Pro sensor. Maximum luminance output of the display monitor was 100 cd/m^2 (50 cd/m^2 mean luminance after calibration), the frame rate was set to 100 Hz, and the resolution was set to 1024×768 pixels. Single pixels subtended 0.0381° of visual angle (i.e., 2.23 arc min.) when viewed from 1.0 m. Head position was maintained using a chin rest, and participant input was recorded via keyboard press.

Stimuli

All stimuli consisted of synthetic visual noise patterns (see Figure 1A) constructed in the Fourier domain using MATLAB (MathWorks, Natick, MA) and corresponding Signal Processing and Image Processing toolboxes. The visual noise stimuli were created by constructing a polar matrix for the amplitude spectrum and assigning all coordinates the same arbitrary amplitude coefficient (except at the location of the DC component, which was assigned a value of 0). The result is a flat isotropic broadband spectrum (i.e., $\alpha = 0.0$), referred to as the template amplitude spectrum (Hansen & Hess, 2006; Tadmor & Tolhurst, 1994). In this form, the α of the template spectrum can be adjusted by multiplying each spatial frequency's amplitude coefficient by $f^{-\alpha}$. The phase spectra were constructed by assigning random values from $-\pi$ to π to the different coordinates of a polar matrix while maintaining an odd-symmetric phase relationship to maintain conjugate symmetry. The noise patterns were rendered into the spatial domain by taking the inverse Fourier transform of an α -altered template amplitude spectrum and a given random phase spectrum. The phase spectrum for all stimuli presented within a trial was identical but randomized from trial-to-trial. We generated $1/f^\alpha$ noise stimuli at

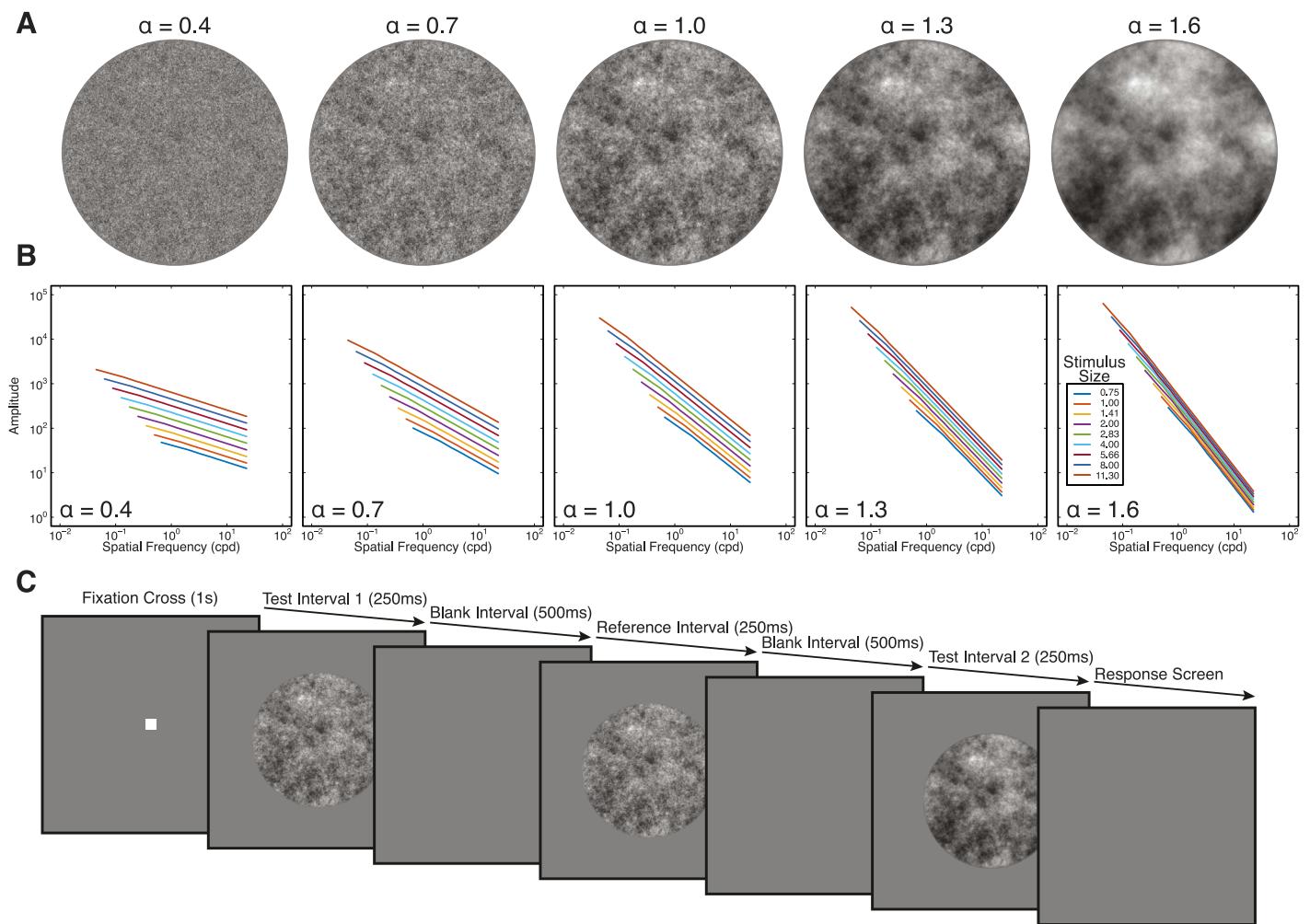


Figure 1. (A) Examples of the five reference α values (0.4, 0.7, 1.0, 1.3, and 1.6) used in this experiment. The phase spectrum of all five stimuli presented is identical. (B) The amplitude spectrum for all five reference α values and nine stimulus sizes presented to observers in this study. Increases in stimulus size lead to additional low spatial frequencies for all stimuli. The smallest stimulus (0.75°) contained nearly four octaves of spatial frequency content ($\min_{sf} = 1.346$ cpd, $\max_{sf} = 22.881$ cpd) while the largest contained approximately eight octaves of spatial frequency content ($\min_{sf} = 0.088$, $\max_{sf} = 22.881$ cpd). (C) The general psychophysical procedure employed in our experiment. Slope (α) discrimination thresholds were estimated with a 3-IFC, 2-AFC “odd-man-out” psychophysical procedure. Observers indicated which, of the first or third interval, was different from the second—reference—interval. The difference in α shown here is exaggerated for print.

nine different diameters³ (0.75° , 1.00° , 1.41° , 2.00° , 2.83° , 4.00° , 5.66° , 8.00° , and 11.30°). RMS contrast (the standard deviation of all pixel luminance values divided by the mean of all pixel luminance values) was fixed to 0.15 for each stimulus size individually. Our RMS contrast calculations for each image were based only on the image region lying inside the circularly windowed area, thus excluding zero-contrast regions.

Psychophysical procedures

The experimental conditions consisted of five reference slope values and nine stimulus sizes for a total of 45 stimulus condition blocks, which were repeated

twice by all observers for a total of 90 blocks. Stimulus size was constant within each block, and the order of blocks was randomized between repetitions and observers. Slope (α) discrimination thresholds were estimated by a temporal three-interval, two-alternative “odd-man-out” forced choice task identical to that used by Hansen and Hess (2006) and Johnson et al. (2011). Participants indicated which, either the first or third stimulus interval, had a different α from the second—reference—stimulus interval (see Figure 1B). At the beginning of each trial, a white (RGB [255, 255, 255]) fixation cross, which subtended 0.3° of visual angle in diameter, was presented for 1 s at the center of the screen. This was followed by three stimulus presentation intervals that each lasted 250 ms and were

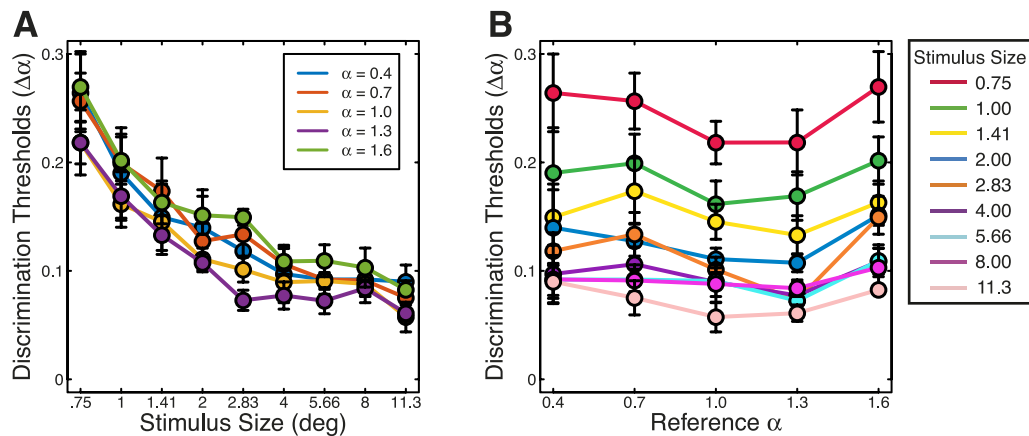


Figure 2. Summary results of the slope discrimination experiment. (A) α discrimination thresholds as a function of stimulus size for each of the five reference α values. As expected, thresholds decreased as a function of an increase in stimulus size. For reference α s of 1.0 and 1.3, thresholds appear to decrease rapidly up to a stimulus size of approximately 2.83° . However, our analyses show no interaction of stimulus size by reference α . (B) The identical data as in (A) but shown with reference α on the x-axis. Each color in this figure corresponds to the reference α (see legend in [A]), while the increase in opacity of the lines marks the increase in stimulus size. Tuning to α was preserved for all stimulus sizes other than 8° . Error bars represent the standard error of the mean.

interlaced by a blank screen (mean luminance) presented for 500 ms. The second interval always contained the reference amplitude spectrum, set at one of the five fixed reference α values (0.4, 0.7, 1.0, 1.3, and 1.6). One interval, either the first or the third, contained the same amplitude spectrum as that of the reference interval, while the other contained the test amplitude spectrum with a steeper α than the α of the reference stimulus. At the end of each trial, the screen was set to mean luminance, and participants indicated which interval, either the first or the third, they perceived as being the odd-man-out via keyboard press. Viewing was binocular and the duration of the response interval was unlimited.

The trial-to-trial change in the image's α was controlled by a 1-up, 2-down staircase procedure using the *PAL_AMUD_setupUD* and *PAL_AMUD_updateUD* functions from the Palamedes toolbox for MATLAB (Prins & Kingdom, 2009). The staircase approached the reference α value from above. The initial α of the odd stimulus was reference $\alpha + 0.5$. The α of the test interval was decreased in linear steps (step size down = 0.02) toward the reference α when the observer made two consecutive correct responses and was increased in linear steps (step size up = 0.02) back toward the start value when the observer made an incorrect response (1-up/2-down rule). The procedure targeted the 70.71% performance level on a psychometric function (Kaernbach, 1991; Prins & Kingdom, 2009). To prevent extreme α values when estimating thresholds, the minimum possible α of the odd stimulus was set to the reference α value and the maximum was set to an $\alpha = 3$. The experimental block continued until 12 reversals had occurred, at which point the block was terminated. Thresholds were estimated by averaging

the α values of the odd stimulus for the last five reversals.

Results

The average effects of stimulus size on α discrimination thresholds ($\Delta\alpha$ s) are shown in Figure 2. Discrimination thresholds were high for small stimuli and fell mostly monotonically as stimulus size increased up to 11.3° of visual angle, $F(8, 48) = 41.85$, $p < 0.001$, $\eta_p^2 = 0.875$. We also find a main effect of reference α on discrimination thresholds, $F(4, 24) = 4.10$, $p = 0.011$, $\eta_p^2 = 0.406$, well-explained by a cubic trend with coefficients $[-1 \ 2 \ 0 \ -2 \ 1]$, $F(1, 6) = 31.16$, $p = 0.001$, $\eta_p^2 = 0.839$. This agrees with previous reports of α discrimination, wherein discrimination thresholds increase from a reference α of 0.4 to 0.7, decrease when the reference α steepens to 1.0 and 1.3, and finally increase for a reference α of 1.6 (Hansen & Hess, 2006; Johnson et al., 2011; Knill et al., 1990; Tadmor & Tolhurst, 1994). The interaction term between stimulus size and reference α was not statistically significant, $F(32, 192) = 0.675$, $p = 0.906$, $\eta_p^2 = 0.101$, and therefore the variation in thresholds across reference α was not affected by stimulus size. It is surprising that tuning to α is preserved across stimulus size here given previous findings that α discrimination thresholds to large stimuli showed no tuning to α (Thomson & Foster, 1997). Notice that discrimination thresholds for the 8° stimulus are flat, showing no specific tuning to α like the other stimulus sizes. This may be attributed to noise in our measurements as the typical tuning function is observed at the larger stimulus size of 11.3° . Preserved

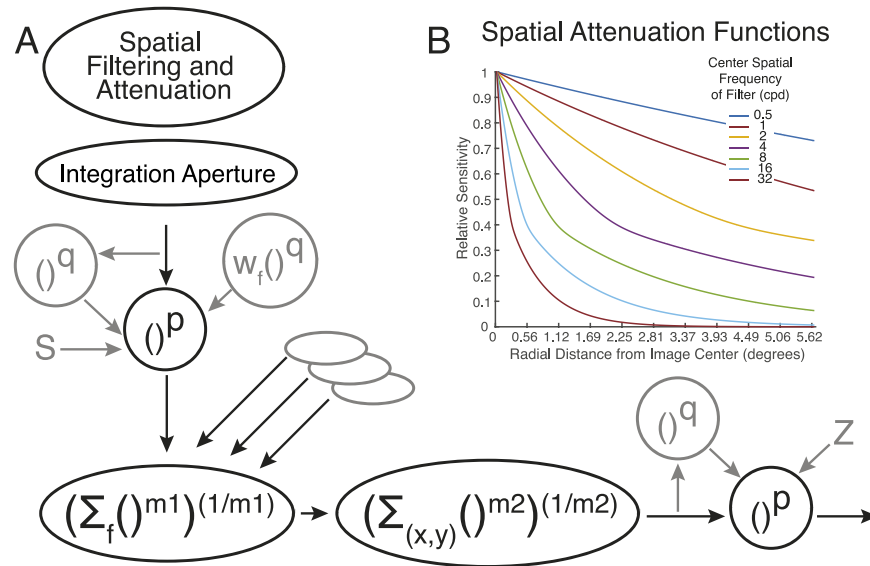


Figure 3. (A) Broadband spatial summation model diagram. The spatially attenuated responses first went through an integration aperture that limited the spatial integration of each spatial frequency channel to 12 cycles of their peak frequency. This was then followed by a contrast gain control operation, which includes a bias in suppression strength towards lower spatial frequency channel responses (w_f). The filter responses are subsequently summed using Minkowski summation, and then summed over space via a second Minkowski summation stage. Finally, the summed output undergoes a second contrast gain control stage (response nonlinearity) prior to the decision stage and discrimination threshold generation. (B) The retinal inhomogeneity function used here to describe the decrease in relative sensitivity for each spatial frequency filter according to the radial distance in degrees. Note that the x-axis marks radial distance from the center of the image in degrees of visual angle but the relative sensitivity of spatial attenuation was calculated in number of cycles of the center spatial frequency of each filter.

tuning is indicative of a mechanism that can maintain sensitivity to α even when low spatial frequency content is added to the image. We explore the form of channel suppression in the modeling section that follows.

Model

Our modeling approach combines previous models of spatial summation and broadband contrast perception (Baker & Meese, 2011; Haun & Peli, 2013; Meese & Baker, 2011). The input to our model were $1/f^\alpha$ noise images at the same resolution as that of our experiment and go through multiple image processing stages (see Figure 3A). Images were spatially filtered by a bank of spatial frequency filters with a bandwidth of ~ 1.5 octaves⁴ (FWHH) and preferred spatial frequencies of 0.5, 1, 2, 4, 8, 16, and 32 c/°. The spatial frequency selectivity of our filters (D) was defined as radial profile of a log Gaussian,

$$D(f) = e^{-\frac{\ln(f/f_0)^2}{2\ln(\sigma/f_0)^2}} \quad (2)$$

where f_0 sets the center spatial frequency of the filter (see above), and the ratio σ/f_0 expresses the spatial frequency bandwidth (FWHH) of our filters ($\sigma/f_0 = 0.65$, ~ 1.5 octaves). Note that because our stimuli are

isotropic (i.e., equal in contrast for all orientations), we do not include any orientation selective filtering in this model. Each spatial frequency filter was adjusted in sensitivity to follow that of a simple approximation of the contrast sensitivity function (CSF) defined as a log-Gaussian ($CS_f = e^{-[(\log_{10}f - \log_{10}f_{\text{peak}})^2/2\sigma_f^2]}$), where f marks the center spatial frequency of the filter, and f_{peak} (1 c/°) is the peak spatial frequency, and σ (1.18) the standard deviation of the CSF (Daly, 1987; Larson & Chandler, 2010). The output of the spatial filtering stage was then multiplied by a spatial attenuation function that describes the decrease in sensitivity that follows the increase in eccentricity from the center of the visual field (see Figure 3B). We use the spatial attenuation function (S), which describe contrast sensitivity across eccentricity, defined by Baldwin, Meese, and Baker (2012) for each spatial frequency filter.

$$S = -\log_{10} \left(\frac{10^{b_1 E}}{10^{(b_1 - b_2)v} + 10^{(b_1 - b_2)E}} \right) + N \quad (3)$$

The parameters b_1 and b_2 control the slope of the functions, E is eccentricity (defined in cycles), v controls the location of the knee point, while N controls the vertical location of the function (the MATLAB code to generate spatial attenuation functions is available online from Baker & Meese, 2011). All parameter values for these functions were fixed and taken from

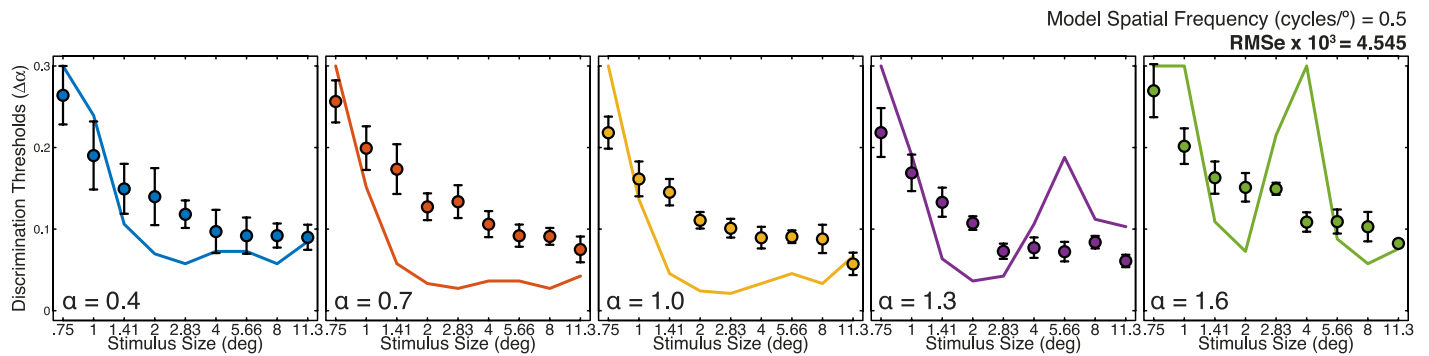


Figure 4. Fits of the single channel model with spatial frequency of 0.5 c/°. Single channel model responses with filters of other peak spatial frequencies are shown in Appendix Figure A3. The data points mark the discrimination thresholds of observers for a given reference α , separated into subplots (reference α is indicated in the bottom left of the subplot). The single channel model outputs are shown as lines.

Baker and Meese (2011). We assume that spatial attenuation is scale invariant, as has been previously proposed (Baker & Meese, 2011; Baldwin et al., 2012; Meese & Baker, 2011). This means that the magnitude of spatial attenuation function for each spatial frequency filter is defined in the number of cycles, which differs according to the peak spatial frequency of the filter, and not degrees of visual angle (see Figure 3B).

Additionally, while observers can integrate contrast over large areas of the visual field, there are nevertheless limits to long-range contrast integration that must be included in our model (Baker & Meese, 2011). We defined the integration aperture as a circular hard-edge window centered on the stimulus. All pixels inside the integration aperture contributed to model output, while pixels outside were discarded. Just as the spatial attenuation function, we defined the size of the integration apertures in number of cycles for each spatial frequency filter (i.e., the integration aperture is scale invariant). We set the size of the integration apertures to 12 cycles, which generated the smallest RMS error in our model simulations (see Figure A1 and A2). This is similar aperture size to those found with previous research on spatial summation with narrowband stimuli (Baker & Meese, 2011).

Single channel model

To use as a reference for the quality of fits of our own model, we first measure how a single channel model performs on our α discrimination threshold data. We took each of the spatially attenuated filter responses (the output of the image processing stages described above) and passed them through a nonlinear transducer (i.e., contrast gain control equation) that included a single self-suppression term unbiased for spatial frequency (Equation 4). This equation is Cannon's spatial model of perceived contrast (Cannon,

1995; Cannon & Fullenkamp, 1991), but we omit summation across spatial frequency channels. The filter responses were then summed over space with Minkowski summation with the exponent taking the value m ,

$$R = \left(\sum_{x,y} \left[\frac{|r(x,y)|^p}{S + |r(x,y)|^q} \right]^m \right)^{\frac{1}{m}}. \quad (4)$$

We set discrimination thresholds for the model when the absolute difference between model responses for the reference α (R_α) and the test α ($R_{\Delta\alpha}$) equaled the sensitivity parameter K ,

$$K = |R_\alpha - R_{\Delta\alpha}| \quad (5)$$

Each spatial frequency model had five free parameters, p , q , m , S , and K . Model fitting was accomplished by optimizing the free parameters with *fminsearch* in MATLAB to minimize the sums of squared error between model output and observer thresholds. The resulting model fits of the single channel with a spatial frequency filter of 0.5 c/° are shown in Figure 4 (the output of models with other spatial frequency can be found in the Appendix, Figure A3). The single channel model with peak spatial frequency of 0.5 c/° captures the general facilitation effects of an increase in stimulus size on discrimination thresholds. However, the single channel model grossly overestimates the magnitude of facilitation at small stimulus sizes, and levels off too quickly (stimulus size of 1.41° for reference α s 0.4–1.0) compared to our observer data. Evidently, the response of a single low spatial frequency channel (or any other single channel spatial frequency) is insufficient to accurately capture the psychophysical performance of our observers. It is more likely, particularly given the nature of our stimuli, that the responses of more than a single spatial frequency channel are contributing to discrimination. We explore how two different multi-channel models would perform in the following section.

One and two stage multichannel models

Two multichannel model variants are tested here: one that includes a single contrast gain control (i.e., a nonlinear transducer) stage, and a second model with an additional contrast gain control stage prior to decision. For both, the first contrast gain control stage (Equation 6) and summation over space (Equation 7) are taken directly from Haun and Peli (2013):

$$\text{resp}(x, y) = \left(\sum_f \left[\frac{|r(x, y)_f|^{p_1}}{S + |r(x, y)_f|^{q_1} + w_f |r(x, y)_f|^{q_1}} \right]^{m_1} \right)^{\frac{1}{m_1}} \quad (6)$$

As in Equation 4, r represents the filter responses of each spatial frequency channel over space and p_1 and q_1 are the excitatory and inhibitory nonlinearities, respectively. We subscript the excitatory and inhibitory exponents here because the second multichannel model includes two stages of contrast gain control with different excitatory and inhibitory exponents. Equation 6 builds on Equation 4 (the single channel model) by adding a second term in the denominator that scales the output of each spatial frequency filter by w_f . There is psychophysical and neurophysiological evidence that low spatial frequency channels receive a disproportionate amount of suppression when responding to broadband, or naturalistic images (Bex et al., 2009; Hansen, Ellemberg, & Johnson, 2012; Hansen, Jacques, Johnson, & Ellemberg, 2011; Haun & Essock, 2010; Meese & Hess, 2004; Webster & Miyahara, 1997). This is implemented in Equation 6 by making w_f inversely proportional to spatial frequency ($1/f$) such that low spatial frequency responses received more divisive inhibition than weaker high spatial frequency responses. Here, we set w_f to $1/f^\beta$ with $\beta = 1.0$, which approximates the average amplitude spectrum of natural scenes (we use β here to identify the model parameter instead of α , the image amplitude spectrum slope; Billock, 2000). Note that we opted to set β to 1.0, but others have selected shallower values in the past that appear to generate reasonable fits as well (Haun & Peli, 2013). The filter responses were then summed across spatial frequency via Minkowski sum set by the exponent m_1 .

The summed channel responses were then combined over space by a second Minkowski sum with the exponent m_2 ,

$$R = \left(\sum_x \sum_y \text{resp}_{x,y}^{m_2} \right)^{\frac{1}{m_2}} \quad (7)$$

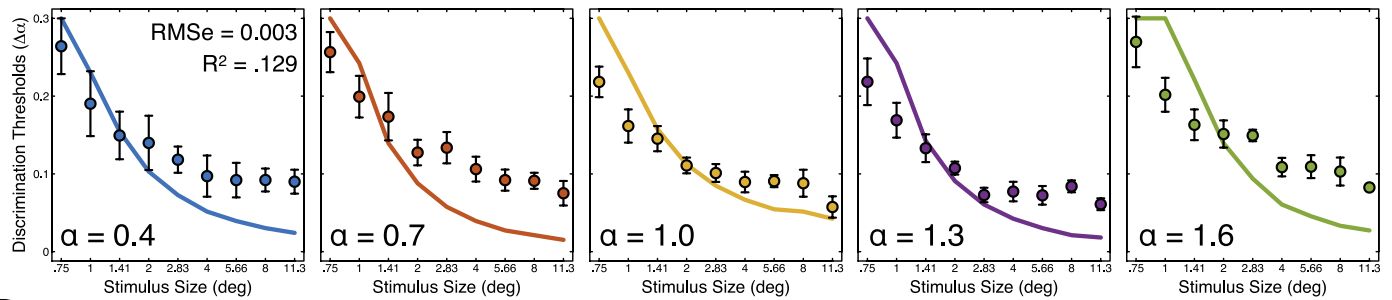
In the single stage model, the output of Equation 7 is fed to the decision stage of the model (Equation 5).

While the sensitivity parameter K was free in our single channel model fitting, it was fixed when fitting the multichannel models. We selected the value of the K parameter by fitting the single stage multi-channel model to our data with K as the only free parameter, all other parameters were fixed to values from other summation models ($p_1 = 2.4$, $q_1 = 2.0$, $m_1 = 4.0$, $m_2 = 4.0$, $S = 1$). The K values were then used as the sensitivity parameter in both the single stage and two stage model variants. Thus, the single stage model had five free parameters (p_1 , q_1 , m_1 , m_2 , S) that were optimized in the same manner as the single channel model. The resulting fit of the single-stage multichannel model of broadband spatial summation is shown in Figure 5A. This model performs better than the single channel models, but is incapable of accurately capturing our data ($r^2 = 0.129$). While all model fits show a decrease in discrimination thresholds according to an increase in stimulus size, all of the models overestimate the summation at large stimulus sizes. Evidently, a model of broadband spatial summation with a single contrast gain control stage is incapable of capturing the effects measured here. However, previous models of suprathreshold spatial summation have included a second non-linear transducer following summation (Meese & Baker, 2011). We decided to include this second contrast gain control stage to our model in order to verify whether it could improve model fits, particularly at larger stimulus sizes. This second contrast gain control term serves to nonlinearly transform the spatially summed output of Equation 7 for input to the decision stage (Baker, Meese, & Georgeson, 2007; Meese, 2010; Meese & Baker, 2011; Meese et al., 2006),

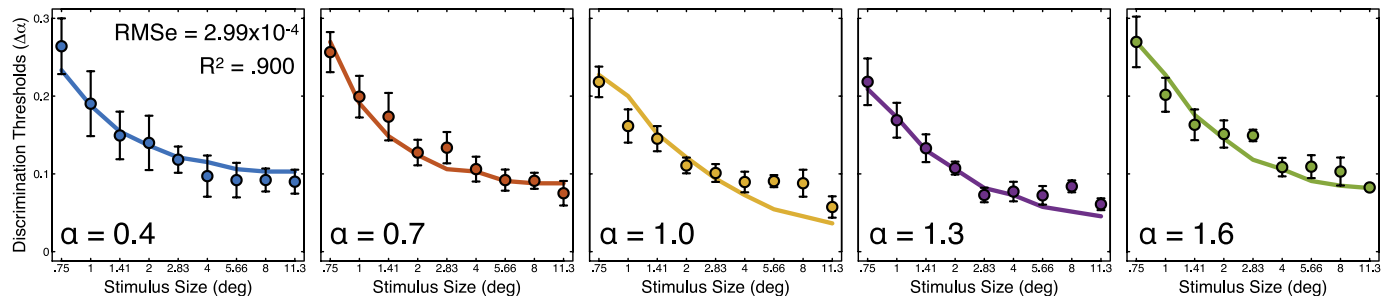
$$R_{\text{final}} = \frac{R^{p_2}}{Z + R^{q_2}} \quad (8)$$

where p_2 and q_2 represent the second excitatory and inhibitory nonlinearities, and R_{final} is the input to the decision stage (Equation 5). The fits of the two-stage model to α discrimination thresholds are shown in Figure 5B. The output nonlinearity to the model improves fits significantly. However, the two-stage model has three additional free parameters (p_2 , q_2 , Z) compared to the single-stage model. The addition of three free parameters is unlikely to explain the improvement in fits between the single and two-stage models. These two models are, however, nested models and we therefore verify that the improvement in model fits were not attributed to the additional free parameters by conducting an extra sums of square F test and calculating AIC scores for each model (Akaike, 1974). The results of both analyses support that the two-stage model ($\text{AIC}_{2\text{-stage}} = -349.11$) is a better descriptor of our data than the single stage model ($\text{AIC}_{1\text{-stage}} = -257.94$) even with the additional free parameters, $F(3,$

A Single Stage Model



B Two-Stage Model (Sum Channels First)



C Two-Stage Model (Sum Space First)

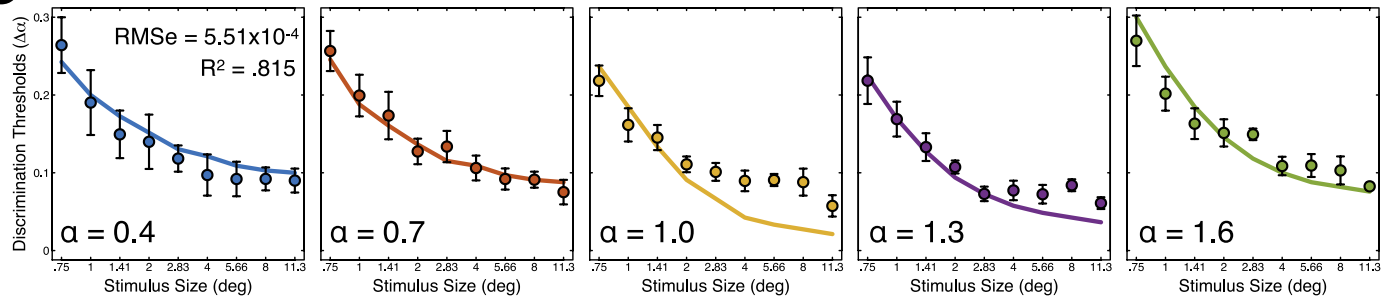


Figure 5. (A) Single stage model fits to observer α discrimination thresholds. Each panel separates α discrimination thresholds by reference α . The circle markers indicate the mean of observer thresholds and the lines are model predictions. The model performs well and captures nearly 86% of the variance in our data. The best fitting parameters of the single stage model were: $p = 2.81$, $q = 2.05$, $m_1 = 4.33$, $m_2 = 4.09$, and $S = 0.98$. (B). Model fits of the two-stage model of broadband spatial summation. The two-stage model explained approximately 91% of the variance in our data. The best fitting parameters of the two-stage model were: $p_1 = 2.54$, $q_1 = 2.18$, $p_2 = 7.03$, $q_2 = 5.93$, $m_1 = 1.11$, $m_2 = 4.65$, $S = 0.98$, and $Z = 0.96$. While the two-stage model has more free parameters than the single stage model, the two-stage model is still a better descriptor of our data ($\Delta AIC_1 = 121.09$ and $\Delta AIC_2 = 0$). (C) Model fits of the two-stage model when the first Minkowski sum is taken over space and over channels second. Best fitting parameters were: $p_1 = 2.75$, $q_1 = 2.30$, $p_2 = 7.01$, $q_2 = 5.65$, $m_1 = 6.27$, $m_2 = 1.81$, $S = 0.99$, and $Z = 0.95$. The different order of operations had a small negative effect on the quality of the fits as it worsened fits for discrimination thresholds with a reference α of 1.0 and 1.3. Note that in both models (B) and (C), the exponent of the Minkowski summation over channels is less than 2, which may indicate near linear summation across channels regardless of the order of operations.

37) = 94.54, $p < 0.001$. The addition of a second nonlinearity appears to be an important addition in characterizing the spatial summation of broadband contrast.

We defined the summation order in our model (summation over channels precedes summation over space) in a manner identical to that of the model of perceived broadband contrast defined by Haun and Peli (2013). Their original model makes no particular claim in regard to the order of summation operations as they

set the Minkowski exponent to be equal for both, and therefore both operations can be collapsed into one. The model parameters we estimate here differ for each summation operation: summation over spatial frequency channels is near linear ($m_1 = 1.11$), while summation over space is closer to standard Minkowski summation ($m_2 = 4.65$). The order of summation operations here is an important factor in the fits of the model, as implementing two Minkowski summations with different exponents in different orders will alter

the results. We chose to verify how summation over space preceding summation over channels might affect model predictions by fitting a model with this summation operation order to our data. Fitting procedures and starting parameters were identical to the original model developed. Fits for the addition model with reversed summation order are shown in Figure 5C. The model captures most of the observer data, but does overestimate the magnitude of spatial summation for reference α s of 1.0 and 1.3 at larger stimulus sizes. This lowered explained variance by nearly 9% compared to the two-stage model where spatial frequency channel responses are summed first. Our original two-stage spatial summation model, where spatial frequency channel responses are summed first, appears to be the best descriptor of our data. Interestingly, in the fitting of both models, the Minkowski exponent for summation over channels remained below a value of 2 (channels first: $m_1 = 1.11$, space first: $m_2 = 1.81$), which is indicative of a quasilinear summation process over channels, irrespective of the order of summation.

Discussion

Decades of studies on spatial vision have generated substantial knowledge of how the early visual system processes luminance contrast (Baker & Meese, 2011; Campbell & Green, 1965; Graham & Robson, 1987; Graham, Robson, & Nachmias, 1978; Graham & Sutter, 1998; Landy & Oruç, 2002; Legge, 1984; Meese & Baker, 2011; Meese & Summers, 2007). These studies have continuously relied on narrowband stimuli to measure the properties of a single channel in isolation. However, the retinal image formed by the visual world is broadband, and therefore multiple channels with different tuning properties will be simultaneously active and interact with each other. Thus, to understand how the visual system operates, it is important to use stimuli that are a better representation of the natural world (Essock et al., 2009; Hansen et al., 2003, 2015; Hansen & Hess, 2012; Haun & Peli, 2013; Meese & Holmes, 2007, 2010; Roeber et al., 2008).

Here, we explored whether discriminability of $1/f^\alpha$ noise broadband stimuli improves as a function of an increase in stimulus size (i.e., broadband spatial summation). We measured discrimination thresholds to broadband stimuli with an α discrimination task that asked of observers to discriminate between noise images that differed in α . As the α of an image represents the distribution of contrast across spatial frequency, α discrimination thresholds can be interpreted as a general measure of broadband contrast sensitivity and thus, serve as a proxy to traditional

measurements of spatial summation. Amplitude spectrum slope discrimination thresholds decreased according to an increase in stimulus size, indicating that broadband stimuli do undergo spatial summation. The decrease in thresholds did not alter tuning to α . Discrimination thresholds were always lowest when the reference α equaled 1.0 or 1.3, regardless of stimulus size. Previous studies on spatial summation at supra-contrast levels have found summation to be best defined as a cascade of operations, whereby the summation stage is preceded and followed by contrast gain control stages (Baker & Meese, 2011; Meese & Baker, 2011). While these studies have been exclusively conducted with narrowband stimuli, the architecture they identified is an excellent starting point from which to develop a model of broadband spatial summation. Indeed, our model required only a few adjustments to account for broadband stimulus input. First, contrast gain control operations were weighted according to spatial frequency in order to apply stronger suppression towards lower spatial frequencies than high. Second, summation over differently tuned spatial frequency channels appears to be near linear while summation over space has a Minkowski exponent of approximately 4, concordant with typical Minkowski summation.

Contrast gain control biases

It is unclear whether or not there is a spatial frequency bias in contrast gain control strength for suprathreshold spatial summation measured with narrowband stimuli. Spatial summation is scale invariant, and thus the magnitude of summation effects is identical across spatial frequency when measured with narrowband stimuli (Baker & Meese, 2011). However, spatial frequency biases in contrast gain control are well-documented in masking studies and suggest stronger divisive inhibition towards low spatial frequencies and high temporal frequency (Cass et al., 2009; Meese & Holmes, 2007, 2010). Similar low spatial frequency biases in contrast gain control should be expected in natural scenes perception given their $1/f^\alpha$ spectra and, indeed, have been observed in measurements of perceived broadband contrast (Haun & Peli, 2013). Given that our stimuli also had $1/f^\alpha$ amplitude spectra, we included a nearly identical low spatial frequency bias in contrast gain control in our model as that defined by Haun and Peli (2013). The low spatial frequency bias was obtained by multiplying the response of each spatial frequency filter by the term $w_f = 1/f^\beta$ with β set to 1.0. This bias in contrast gain control is steeper than that used by Haun and Peli (2013). We did not attempt to fit the exact value of β in this study but did find that β s set to be shallower than

1.0 generated worse fits than when the β of the weighting term was 1.0 or steeper: varying β between 1.0 and 1.6 (the steepest reference α used here for our stimuli) did not significantly impact the goodness-of-fit of the model. We selected a β set to 1.0 in our simulation as this value is closest to the average α of natural scenes, and best represents the typical input received by the visual system (Billock, 2000; Hansen & Essock, 2005; Tolhurst et al., 1992), but other, steeper, values may also be appropriate.

There are additional aspects of bias suppression in contrast gain control that we do not estimate in our modeling and may limit our ability to measure the exact form any spatial frequency bias. For example, we do not attempt to define any suppressive interactions between orientation tuned channels (Foley, 1994; Meese & Holmes, 2007, 2010), distant spatial frequency channels (Foley, 1994) or different spatial locations (i.e., lateral suppression; Cannon & Fullenkamp, 1991; Chen & Tyler, 2001, 2008; Meese, 2004; Xing & Heeger, 2000). A dataset that explicitly sets out to describe these additional suppressive processes may be better suited to characterize biases in suppression, but this is beyond the scope of the current study.

Linearity and order of summation operations

Models of spatial summation developed with narrowband stimuli have found the processing of summation to be best defined as a cascade of nonlinear transducers (e.g., contrast gain control) and summation stages (Baker & Meese, 2011; Foley et al., 2007; Meese, 2004; Meese & Baker, 2011; Meese & Summers, 2007; Wilson & Gelb, 1984). While details vary across models, the most recent implementation found summation to be linear within the integration aperture (pixel-wise summation), while summation across apertures is nonlinear. These models were not designed to account for broadband stimulus input. They can be adapted to do so by including an additional suppressive term in the contrast gain control operation that may or may not bias suppression of spatial frequency channels, as has been demonstrated here and elsewhere (Alam et al., 2014; Chandler, Gaubatz, & Hemami, 2009; Haun & Peli, 2013; Huang & Dai, 2018). That said, our model is not a simple replication of narrowband models of spatial summation, as we have to account for the integration of differently tuned spatial frequency channel responses. Evidently, models of spatial summation built with narrowband stimuli cannot comment on the order of operations between summation over channels (either spatial frequency or orientation) and summation over space. Previous models of broadband contrast perception either make no assumption on the order of summation operations (by placing the same

Minkowski exponent for both; Haun & Peli, 2013) or place summation over channels prior to summation over space (Alam et al., 2014; Chandler et al., 2009). We found the best fitting model to sum over channels prior to summing over space. This means that summation is late in our processing stream, and potentially differs from narrowband models that will sum over small regions of space early on (e.g., Baker & Meese, 2011).

We also found that summation across spatial frequency channels in our model is near linear. The Minkowski exponent for summation across channels was 1.11, which suggests that all channels contribute relatively equally to spatial summation. Other models that include summation across multiple, differently tuned, spatial frequency channels have typically used Minkowski summation exponents of 3.0 or 4.0, which is intended to bias the response towards a spatial frequency specific, or winner-take-all, response (Cannon, 1995; Haun & Peli, 2013). Linear summation is not, however, completely unexpected. There are many spatial summation or binocular summation models that implement linear summation across narrow regions of space or across eyes (Baker & Meese, 2011; Foley et al., 2007; Huang & Dai, 2018; Meese, 2004; Meese & Baker, 2011; Meese & Summers, 2007). There are also models of broadband visual masking that use Minkowski exponents of less than 2.0 when summing over visual channels (orientation and spatial frequency; Alam et al., 2014; Chandler et al., 2009). We should note that while the summation process appears to be near linear, the input to the summation is not. The spatial frequency channel responses are rectified and go through a nonlinear transducer (contrast gain control) prior to summation. Nevertheless, the finding that the summation of spatial frequency channel responses is near-linear is interesting as it suggests that the responses of all spatial frequency channels are important in the discrimination of broadband contrast.

Sensitivity to α

We were surprised to observe that tuning to α was preserved across all stimulus sizes used in this study. Lower discrimination thresholds for reference α s between 1.0–1.3 is common when discrimination thresholds are measured with smaller stimuli (1° – 2° ; Hansen & Hess, 2006; Knill et al., 1990). However, the only study that used large images to measure α discrimination thresholds (Thomson & Foster, 1997) found no indication of tuning to α s between 1.0 and 1.3. There are methodological differences between the current study and that of Thomson and Foster (1997) that may account for the discrepancy in results. To measure α discrimination thresholds, we opted to

generate synthetic broadband noise stimuli that offer complete control over the construction of the amplitude and phase spectra. These noise images are naturalistic only in their amplitude spectrum but are a practical choice because they simplify the investigation of the combination of contrast across spatial frequency and space, and control for any additional factors that may influence α discrimination. One such factor is the uneven distribution of luminance contrast across orientation in natural scenes, with more contrast present at cardinals than obliques (Betsch, Einhäuser, Körding, & König, 2004; Essock, DeFord, Hansen, & Sinai, 2003; Hansen & Essock, 2006; Hansen et al., 2003). Thomson and Foster (1997), however, used natural and phase randomized natural scenes and physically modulated the α of these images, which preserves biases in contrast across orientation even for phase scrambled images. Humans are not equally sensitive to orientation and in broadband images as they have worse sensitivity to cardinal than oblique orientations (Essock et al., 2003; Hansen et al., 2015; Hansen, Haun, & Essock, 2008; Hansen & Essock, 2006; Haun & Essock, 2010). When measuring α discrimination thresholds with natural scenes, it is likely that any biases in sensitivity to oriented contrast will influence the ability of observers to correctly complete the α discrimination task and consequently alter tuning to α and may account for any discrepancy between our findings and those of Thomson and Foster (1997).

Conclusion

We verified if broadband stimuli that contain the characteristic $1/f^\alpha$ amplitude spectrum of natural scenes are subject to spatial summation (i.e., increase in sensitivity according to an increase in stimulus size) and if psychophysical models of summation developed for narrowband stimuli could be adapted to describe broadband spatial summation. We found broadband stimuli are subject to spatial summation: discrimination thresholds were inversely related to stimulus size. We found that a model of spatial summation developed with narrowband stimuli (Baker & Meese, 2011; Meese & Baker, 2011) was capable of capturing our broadband summation effects when modified to account for the broadband aspects of our stimuli. Our data were best fit by a two-stage contrast gain control summation model where spatial frequency channel responses undergo a bias contrast gain control operation that suppresses low spatial frequency channel responses more than high. Summation over spatial frequency channels was near linear and preceded summation over space, which was nonlinear with a Minkowski exponent of approximately 4.0. Our efforts in this study were

concentrated on the processing of spatial frequency content and omitted other aspects of natural scenes (e.g., the distribution of contrast across orientation) that are relevant to perception. How psychophysical models of spatial vision may be further adapted to incorporate additional components of natural scenes remains to be determined. Nevertheless, our findings here are a demonstration that narrowband psychophysical models can serve as adequate starting points to develop models of broadband contrast perception.

Keywords: spatial summation, amplitude spectrum slope, contrast gain control, nonlinear summation, natural image, broadband contrast, spatial frequency channels, contrast integration, computational modeling

Acknowledgments

This work was funded in part by NSF Grant CHS-1524888 to PS; a James S. McDonnell Foundation Grant (220020430) to BCH; a Natural Sciences and Engineering Research Council of Canada grant to AJ; and a Fonds de Recherche du Québec—Nature et technologies graduate fellowship to BR. We are grateful to Daniel H. Baker for his comments and feedback on this manuscript.

Commercial relationships: none.

Corresponding author: Bruno Richard.

Email: bruno.richard@rutgers.edu.

Address: Department of Mathematics and Computer Science, Rutgers University, Newark, NJ, USA.

Footnotes

¹ We define the spatial frequency content here in cycles per degree of visual angle, and an increase in stimulus size as an increase of a window size on a stimulus background. In this context, the resolution upper limit is set by the pixel density of the stimulus and is unchanging across stimulus area.

² Note that we only used 256 gray levels from the palette of 1,024 (pixel intensity was defined in the range of [0–255]).

³ An increase in the area of a stimulus may affect summation differently across stimulus factors because of factors like retinal inhomogeneity (Baker & Meese, 2011; Hess & Hayes, 1994; Kelly, 1984; Meese & Baker, 2011). This has previously motivated the use of other methods that keep stimulus diameter fixed (e.g., “Swiss-cheese” stimuli; Baker & Meese, 2011; Meese & Baker, 2011). These methods are inappropriate for our stimulus type. The Swiss-cheese method, for example,

which manipulates area of a sine-wave carrier by a raised plaid pattern, will introduce sidebands around the carrier frequency, and for a broadband noise image, will alter the spatial frequency spectra of the stimulus, which is explicitly something we want to avoid. When measuring α discrimination thresholds, it is important for the entire spatial frequency spectrum of the image to be unaltered for observer thresholds to show peak sensitivities at α s of 1.0–1.3. Observer thresholds differ significantly from this typical tuning when segments of spatial frequency content are removed (Richard, Hansen, Ellemberg, & Johnson, 2013).

⁴ Note that there is little empirical evidence to support constant bandwidth spatial frequency channels in the human visual system, as most evidence indicates that the bandwidth of spatial frequency channels decreases with increasing spatial frequency (e.g., Wilson, McFarlane, & Phillips, 1983). However, as this study is a preliminary step in computationally describing the visual discrimination of broadband contrast, we opted to simplify our spatial frequency filter bank by using constant 1.5 octave bandwidth filters.

References

- Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control*, 19(6), 716–723. <https://doi.org/10.1109/TAC.1974.1100705>.
- Alam, M. M., Vilankar, K. P., Field, D. J., & Chandler, D. M. (2014). Local masking in natural images: A database and analysis. *Journal of Vision*, 14(8):22, 1–38. <https://doi.org/10.1167/14.8.22>. [PubMed] [Article]
- Baker, D. H., & Meese, T. S. (2011). Contrast integration over area is extensive: A three-stage model of spatial summation. *Journal of Vision*, 11(14):14, 1–16. <https://doi.org/10.1167/11.14.14>. [PubMed] [Article]
- Baker, D. H., Meese, T. S., & Georgeson, M. A. (2007). Binocular interaction: Contrast matching and contrast discrimination are predicted by the same model. *Spatial Vision*, 20(5), 397–413. <https://doi.org/10.1163/156856807781503622>.
- Baldwin, A. S., Meese, T. S., & Baker, D. H. (2012). The attenuation surface for contrast sensitivity has the form of a witch's hat within the central visual field. *Journal of Vision*, 12(11):23, 1–17. <https://doi.org/10.1167/12.11.23>. [PubMed] [Article]
- Betsch, B. Y., Einhäuser, W., Körding, K. P., & König, P. (2004). The world from a cat's perspective—Statistics of natural videos. *Biological Cybernetics*, 90(1), 41–50. <https://doi.org/10.1007/s00422-003-0434-6>.
- Bex, P. J., Mareschal, I., & Dakin, S. C. (2007). Contrast gain control in natural scenes. *Journal of Vision*, 7(11):12, 1–12. <https://doi.org/10.1167/7.11.12>. [PubMed] [Article]
- Bex, P. J., Solomon, S. G., & Dakin, S. C. (2009). Contrast sensitivity in natural scenes depends on edge as well as spatial frequency structure. *Journal of Vision*, 9(10):1, 1–19. <https://doi.org/10.1167/9.10.1>. [PubMed] [Article]
- Billock, V. (2000). Neural acclimation to 1/f spatial frequency spectra in natural images transduced by the human visual system. *Physica D: Nonlinear Phenomena*, 137(3–4), 379–391. [https://doi.org/10.1016/S0167-2789\(99\)00197-9](https://doi.org/10.1016/S0167-2789(99)00197-9).
- Burton, G. J., & Moorhead, I. R. (1987). Color and spatial structure in natural scenes. *Applied Optics*, 26(1), 157–170. <https://doi.org/10.1364/AO.26.000157>.
- Campbell, F. W., & Green, D. G. (1965, October 9). Monocular versus binocular visual acuity. *Nature*, 208(5006), 191–192. <https://doi.org/10.1038/208191a0>.
- Cannon, M. W. (1995). A multiple spatial filter model for suprathreshold contrast perception. In E. Peli (Ed.), *Vision models for target detection and recognition* (pp. 88–116). Singapore: World Scientific.
- Cannon, M. W., & Fullenkamp, S. C. (1991). A transducer model for contrast perception. *Vision Research*, 31(6), 983–998. [https://doi.org/10.1016/S0042-6989\(05\)80001-X](https://doi.org/10.1016/S0042-6989(05)80001-X).
- Cass, J., Stuit, S., Bex, P., & Alais, D. (2009). Orientation bandwidths are invariant across spatiotemporal frequency after isotropic components are removed. *Journal of Vision*, 9(12):17, 1–14. <https://doi.org/10.1167/9.12.17>. [PubMed] [Article]
- Chandler, D. M., Gaubatz, M. D., & Hemami, S. S. (2009). A patch-based structural masking model with an application to compression. *Journal on Image and Video Processing*, 5, 1–22. doi:10.1155/2009/649316.
- Chen, C. C., & Tyler, C. W. (2001). Lateral sensitivity modulation explains the flanker effect in contrast discrimination. *Proceedings. Biological Sciences/The Royal Society*, 268(1466), 509–516. <https://doi.org/10.1098/rspb.2000.1387>.
- Chen, C. C., & Tyler, C. W. (2008). Excitatory and inhibitory interaction fields of flankers revealed by contrast-masking functions. *Journal of Vision*, 8(4):10, 1–14. <https://doi.org/10.1167/8.4.10>. [PubMed] [Article]

- Cunningham, D. G. M., Baker, D. H., & Peirce, J. W. (2017). Measuring nonlinear signal combination using EEG. *Journal of Vision*, 17(5):10, 1–14. <https://doi.org/10.1167/17.5.10>. [PubMed] [Article]
- Daly, S. (1987). *Subroutine for the generation of a two dimensional human visual contrast sensitivity function* (Technical Report Y, 233203). Rochester, NY: Eastman Kodak.
- Ellemberg, D., Hansen, B. C., & Johnson, A. (2012). The developing visual system is not optimally sensitive to the spatial statistics of natural images. *Vision Research*, 67, 1–7. <https://doi.org/10.1016/j.visres.2012.06.018>.
- Essock, E. A., DeFord, J. K., Hansen, B. C., & Sinai, M. J. (2003). Oblique stimuli are seen best (not worst!) in naturalistic broad-band stimuli: A horizontal effect. *Vision Research*, 43(12), 1329–1335. [https://doi.org/10.1016/S0042-6989\(03\)00142-1](https://doi.org/10.1016/S0042-6989(03)00142-1).
- Essock, E. A., Haun, A. M., & Kim, Y. J. (2009). An anisotropy of orientation-tuned suppression that matches the anisotropy of typical natural scenes. *Journal of Vision*, 9(1):35, 1–15. <https://doi.org/10.1167/9.1.35>. [PubMed] [Article]
- Field, D. J. (1987). Relations between the statistics of natural images and the response properties of cortical cells. *Journal of the Optical Society of America A*, 4(12), 2379–2394. <https://doi.org/10.1364/JOSAA.4.002379>.
- Foley, J. M. (1994). Human luminance pattern-vision mechanisms: Masking experiments require a new model. *Journal of the Optical Society of America A*, 11(6), 1710–1719. <https://doi.org/10.1364/JOSAA.11.001710>.
- Foley, J. M., Varadharajan, S., Koh, C. C., & Farias, M. C. Q. (2007). Detection of Gabor patterns of different sizes, shapes, phases and eccentricities. *Vision Research*, 47(1), 85–107. <https://doi.org/10.1016/j.visres.2006.09.005>.
- Georgeson, M. A., Wallis, S. A., Meese, T. S., & Baker, D. H. (2016). Contrast and lustre: A model that accounts for eleven different forms of contrast discrimination in binocular vision. *Vision Research*, 129, 98–118. <https://doi.org/10.1016/j.visres.2016.08.001>.
- Graham, N., & Robson, J. G. (1987). Summation of very close spatial frequencies: The importance of spatial probability summation. *Vision Research*, 27(11), 1997–2007.
- Graham, N., & Sutter, A. (1998). Spatial summation in simple (Fourier) and complex (non-Fourier) texture channels. *Vision Research*, 38(2), 231–257. [https://doi.org/10.1016/S0042-6989\(97\)00154-5](https://doi.org/10.1016/S0042-6989(97)00154-5).
- Graham, N. V. (1977). Visual detection of aperiodic spatial stimuli by probability summation among narrowband channels. *Vision Research*, 17, 637–652.
- Graham, N. V., Robson, J. G., & Nachmias, J. (1978). Grating summation in fovea and periphery. *Vision Research*, 18(7), 815–825. [https://doi.org/10.1016/0042-6989\(78\)90122-0](https://doi.org/10.1016/0042-6989(78)90122-0).
- Graham, N. V., & Sutter, A. (2000). Normalization: Contrast-gain control in simple (Fourier) and complex (non-Fourier) pathways of pattern vision. *Vision Research*, 40(20), 2737–2761. [https://doi.org/10.1016/S0042-6989\(00\)00123-1](https://doi.org/10.1016/S0042-6989(00)00123-1).
- Hansen, B. C., Ellemberg, D., & Johnson, A. P. (2012). Different spatial frequency bands selectively signal for natural image statistics in the early visual system. *Journal of Neurophysiology*, 108(8), 2160–2172. <https://doi.org/10.1152/jn.00288.2012>.
- Hansen, B. C., & Essock, E. A. (2005). Influence of scale and orientation on the visual perception of natural scenes. *Visual Cognition*, 12(6), 1199–1234. <https://doi.org/10.1080/13506280444000715>.
- Hansen, B. C., & Essock, E. A. (2006). Anisotropic local contrast normalization: The role of stimulus orientation and spatial frequency bandwidths in the oblique and horizontal effect perceptual anisotropies. *Vision Research*, 46(26), 4398–4415. <https://doi.org/10.1016/j.visres.2006.07.016>.
- Hansen, B. C., Essock, E. A., Zheng, Y., & DeFord, J. K. (2003). Perceptual anisotropies in visual processing and their relation to natural image statistics. *Network: Computation in Neural Systems*, 14(3), 501–526. https://doi.org/10.1088/0954-898X_14_3_307.
- Hansen, B. C., Haun, A. M., & Essock, E. A. (2008). The horizontal effect: A perceptual anisotropy in visual processing of naturalistic broadband stimuli. In T. A. Portocello & R. V. Velloti (Eds.), *Visual cortex: New research* (pp. 1–34). Hauppauge, NY: Nova Science Publishers.
- Hansen, B. C., & Hess, R. F. (2006). Discrimination of amplitude spectrum slope in the fovea and para-fovea and the local amplitude distributions of natural scene imagery. *Journal of Vision*, 6(7):3, 696–711. <https://doi.org/10.1167/6.7.3>. [PubMed] [Article]
- Hansen, B. C., & Hess, R. F. (2012). On the effectiveness of noise masks: Naturalistic vs. unnaturalistic image statistics. *Vision Research*, 60, 101–113. <https://doi.org/10.1016/j.visres.2012.03.017>.
- Hansen, B. C., Jacques, T., Johnson, A. P., & Ellemberg, D. (2011). From spatial frequency

- contrast to edge preponderance: The differential modulation of early visual evoked potentials by natural scene stimuli. *Visual Neuroscience*, 28(3), 221–237. <https://doi.org/10.1017/S095252381100006X>.
- Hansen, B. C., Richard, B., Andres, K., Johnson, A. P., Thompson, B., & Essock, E. A. (2015). A cortical locus for anisotropic overlay suppression of stimuli presented at fixation. *Visual Neuroscience*, 32, E023. <https://doi.org/10.1017/S0952523815000255>.
- Haun, A. M., & Essock, E. A. (2010). Contrast sensitivity for oriented patterns in 1/f noise: Contrast response and the horizontal effect. *Journal of Vision*, 10(10):1, 1–21, <https://doi.org/10.1167/10.10.1>. [PubMed] [Article]
- Haun, A. M., & Peli, E. (2013). Perceived contrast in complex images. *Journal of Vision*, 13(13):3, 1–21, <https://doi.org/10.1167/13.13.3>. [PubMed] [Article]
- Hess, R. F., & Hayes, A. (1994). The coding of spatial position by the human visual system: Effects of spatial scale and retinal eccentricity. *Vision Research*, 34(5), 625–643. [https://doi.org/10.1016/0042-6989\(94\)90018-3](https://doi.org/10.1016/0042-6989(94)90018-3).
- Huang, P.-C., & Dai, Y.-M. (2018). Binocular contrast-gain control for natural scenes: Image structure and phase alignment. *Vision Research*, 146–147(April), 18–31. <https://doi.org/10.1016/j.visres.2018.02.012>.
- Johnson, A. P., Richard, B., Hansen, B. C., & Ellemberg, D. (2011). The magnitude of center-surround facilitation in the discrimination of amplitude spectrum is dependent on the amplitude of the surround. *Journal of Vision*, 11(7):14, 1–10, <https://doi.org/10.1167/11.7.14>. [PubMed] [Article]
- Kaernbach, C. (1991). Simple adaptive testing with the weighted up-down method. *Perception & Psychophysics*, 49(3), 227–229. <https://doi.org/10.3758/BF03214307>.
- Kelly, D. H. (1984). Retinal inhomogeneity. I. Spatio-temporal contrast sensitivity. *Journal of the Optical Society of America. A, Optics and Image Science*, 1(1), 107–113. <https://doi.org/10.1364/JOSAA.1.000107>.
- Kersten, D. (1984). Spatial summation in visual noise. *Vision Research*, 24(12), 1977–1990. [https://doi.org/10.1016/0042-6989\(84\)90033-6](https://doi.org/10.1016/0042-6989(84)90033-6).
- Knill, D. C., Field, D. J., & Kersten, D. (1990). Human discrimination of fractal images. *Journal of the Optical Society of America. A, Optics and Image Science*, 7(6), 1113–1123. <https://doi.org/10.1364/JOSAA.7.001113>.
- Landy, M. S., & Oruç, I. (2002). Properties of second-order spatial frequency channels. *Vision Research*, 42(19), 2311–2329. [https://doi.org/10.1016/S0042-6989\(02\)00193-1](https://doi.org/10.1016/S0042-6989(02)00193-1).
- Larson, E. C., & Chandler, D. M. (2010). Most apparent distortion: Full-reference image quality assessment and the role of strategy. *Journal of Electronic Imaging*, 19(1):011006, 1–20. <https://doi.org/10.1117/1.3267105>.
- Legge, G. E. (1984). Binocular contrast summation—II. Quadratic summation. *Vision Research*, 24(4), 385–394. [https://doi.org/10.1016/0042-6989\(84\)90064-6](https://doi.org/10.1016/0042-6989(84)90064-6).
- Legge, G. E., & Foley, J. M. (1980). Contrast masking in human vision. *JOSA*, 70(12), 1458–1471. <https://doi.org/10.1364/JOSA.70.001458>.
- Meese, T. S. (2004). Area summation and masking. *Journal of Vision*, 4(10):8, 930–943, <https://doi.org/10.1167/4.10.8>. [PubMed] [Article]
- Meese, T. S. (2010). Spatially extensive summation of contrast energy is revealed by contrast detection of micro-pattern textures. *Journal of Vision*, 10(8):14, 1–21, <https://doi.org/10.1167/10.8.14>. [PubMed] [Article]
- Meese, T. S., & Baker, D. H. (2011). Contrast summation across eyes and space is revealed along the entire dipper function by a “Swiss cheese” stimulus. *Journal of Vision*, 11(1):23, 1–23, <https://doi.org/10.1167/11.1.23>. [PubMed] [Article]
- Meese, T. S., Georgeson, M. A., & Baker, D. H. (2006). Binocular contrast vision at and above threshold. *Journal of Vision*, 6(11):7, 1224–1243, <https://doi.org/10.1167/6.11.7>. [PubMed] [Article]
- Meese, T. S., & Hess, R. F. (2004). Low spatial frequencies are suppressively masked across spatial scale, orientation, field position, and eye of origin. *Journal of Vision*, 4(10):2, 843–859, <https://doi.org/10.1167/4.10.2>. [PubMed] [Article]
- Meese, T. S., Hess, R. F., & Williams, C. B. (2005). Size matters, but not for everyone: Individual differences for contrast discrimination. *Journal of Vision*, 5(11):2, 928–947, <https://doi.org/10.1167/5.11.2>. [PubMed] [Article]
- Meese, T. S., & Holmes, D. J. (2007). Spatial and temporal dependencies of cross-orientation suppression in human vision. *Proceedings. Biological Sciences, The Royal Society*, 274(1606), 127–136. <https://doi.org/10.1098/rspb.2006.3697>.
- Meese, T. S., & Holmes, D. J. (2010). Orientation masking and cross-orientation suppression (XOS): Implications for estimates of filter bandwidth. *Journal of Vision*, 10(12):9, 1–20, <https://doi.org/10.1167/10.12.9>. [PubMed] [Article]
- Meese, T. S., & Summers, R. J. (2007). Area

- summation in human vision at and above detection threshold. *Proceedings. Biological Sciences, The Royal Society*, 274(September), 2891–2900. <https://doi.org/10.1098/rspb.2008.3002>.
- Petrov, Y., Carandini, M., & McKee, S. (2005). Two distinct mechanisms of suppression in human vision. *The Journal of Neuroscience: The Official Journal of the Society for Neuroscience*, 25(38), 8704–8707. <https://doi.org/10.1523/JNEUROSCI.2871-05.2005>.
- Prins, N., & Kingdom, F. A. A. (2009). *Palamedes: Matlab routines for analyzing psychophysical data*. Retrieved from <http://www.palamedestoolbox.org/>.
- Richard, B., Chadnova, E., & Baker, D. H. (2018). Binocular vision adaptively suppresses delayed monocular signals. *NeuroImage*, 172, 753–765. <https://doi.org/10.1016/j.neuroimage.2018.02.021>.
- Richard, B., Hansen, B. C., Ellemberg, D., & Johnson, A. P. (2013). Size dependent increase in sensitivity to the slope of the amplitude spectrum is not solely dependent on the increased low spatial frequency representation of larger stimuli. *Journal of Vision*, 13(9): 1238–1238, <https://doi.org/10.1167/13.9.1238>. [Abstract]
- Robson, J. G., & Graham, N. (1981). Probability summation and regional variation in contrast sensitivity across the visual field. *Vision Research*, 21(3), 409–418. [https://doi.org/10.1016/0042-6989\(81\)90169-3](https://doi.org/10.1016/0042-6989(81)90169-3).
- Roeber, U., Wong, E. M. Y., & Freeman, A. W. (2008). Cross-orientation interactions in human vision. *Journal of Vision*, 8(3):15, 1–11, <https://doi.org/10.1167/8.3.15>. [PubMed] [Article]
- Schwartz, O., & Simoncelli, E. P. (2001). Natural signal statistics and sensory gain control. *Nature Neuroscience*, 4(8), 819–825. <https://doi.org/10.1038/90526>.
- Summers, R. J., Baker, D. H., & Meese, T. S. (2015). Area summation of first- and second-order modulations of luminance. *Journal of Vision*, 15(1):12, 1–13, <https://doi.org/10.1167/15.1.12>. [PubMed] [Article]
- Tadmor, Y., & Tolhurst, D. J. (1994). Discrimination of changes in the second-order statistics of natural and synthetic images. *Vision Research*, 34(4), 541–554. [https://doi.org/10.1016/0042-6989\(94\)90167-8](https://doi.org/10.1016/0042-6989(94)90167-8).
- Thomson, M. G. A., & Foster, D. H. (1997). Role of second- and third-order statistics in the discriminability of natural images. *Journal of the Optical Society of America A*, 14(9), 2081–2092. <https://doi.org/10.1364/JOSAA.14.002081>.
- Tolhurst, D. J., & Tadmor, Y. (1997). Band-limited contrast in natural images explains the detectability of changes in the amplitude spectra. *Vision Research*, 37(23), 3203–3215. [https://doi.org/10.1016/S0042-6989\(97\)00119-3](https://doi.org/10.1016/S0042-6989(97)00119-3).
- Tolhurst, D. J., & Tadmor, Y. (1997). Discrimination of changes in the slopes of the amplitude spectra of natural images: Band-limited contrast and psychometric functions. *Perception*, 26(8), 1011–1025. <https://doi.org/10.1068/p261011>.
- Tolhurst, D. J., Tadmor, Y., & Chao, T. (1992). Amplitude spectra of natural images. *Ophthalmic and Physiological Optics*, 12(2), 229–232. <https://doi.org/10.1111/j.1475-1313.1992.tb00296.x>.
- van der Schaaf, A., & van Hateren, J. H. (1996). Modelling the power spectra of natural images: Statistics and information. *Vision Research*, 36(17), 2759–2770. [https://doi.org/10.1016/0042-6989\(96\)00002-8](https://doi.org/10.1016/0042-6989(96)00002-8).
- Webster, M. A., & Miyahara, E. (1997). Contrast adaptation and the spatial structure of natural images. *Journal of the Optical Society of America A*, 14(9), 2355–2366. <https://doi.org/10.1364/JOSAA.14.002355>.
- Wilson, H. R., McFarlane, D. K., & Phillips, G. C. (1983). Spatial frequency tuning of orientation selective units estimated by oblique masking. *Vision Research*, 23(9), 873–882. [https://doi.org/10.1016/0042-6989\(83\)90055-X](https://doi.org/10.1016/0042-6989(83)90055-X).
- Wilson, H. R., & Gelb, D. J. (1984). Modified line-element theory for spatial-frequency and width discrimination. *Journal of the Optical Society of America A*, 1(1), 124–131. <https://doi.org/10.1364/JOSAA.1.000124>.
- Xing, J., & Heeger, D. J. (2000). Center-surround interactions in foveal and peripheral vision. *Vision Research*, 40(22), 3065–3072. [https://doi.org/10.1016/S0042-6989\(00\)00152-8](https://doi.org/10.1016/S0042-6989(00)00152-8).

Appendix

Integration aperture radius

The radius of the integration aperture (12 cycles) was chosen by fitting the two-stage models (summation over channels first) with different aperture sizes (from 4 to 64 cycles) and selecting that which generated the smallest RMSe (see Figure A1 and A2). Our results resemble those of Baker and Meese (2011), who found that contrast integration extends for at least eight cycles, with an optimal summation region of approximately 12 cycles. However, unlike their findings, larger integration apertures in our model resulted in poorer fits to our data than smaller apertures.

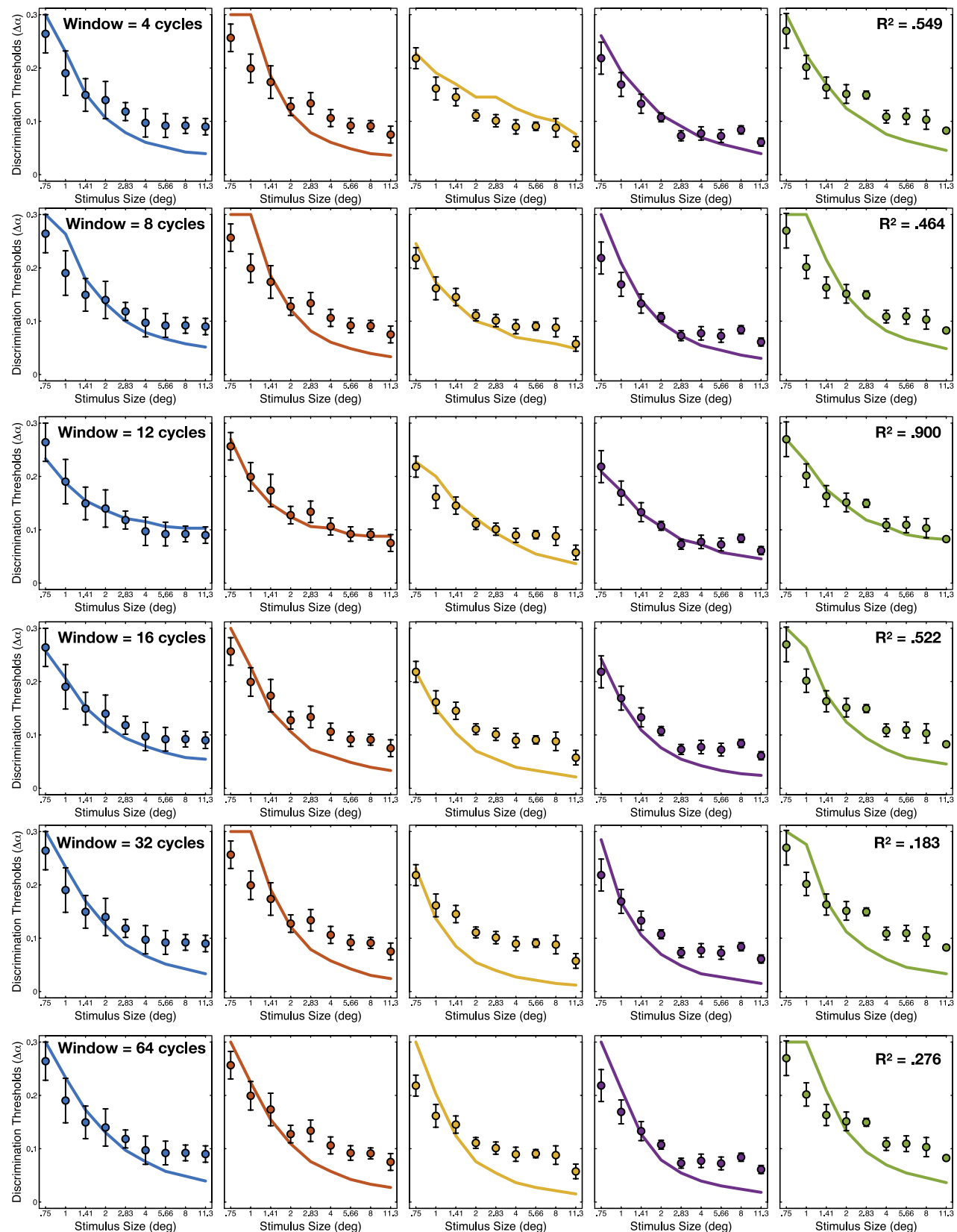


Figure A1. Integration aperture size effects on the fits of both the single stage (dashed line) and two-stage (solid line) models of broadband spatial summation for each reference α . RMSe for both models was smallest for an aperture size of 12 cycles (see Figure A2). For smaller stimulus sizes, models underestimated thresholds for shallow reference α s for small stimulus sizes, while the two-stage model was capable of reaching human observer thresholds for larger stimuli. For steeper reference α s, both models overestimated thresholds for smaller stimulus sizes. As the aperture size increased, both models underestimated thresholds of large stimulus sizes for steep reference α s. Surprisingly, we found little change in thresholds for shallow reference α s at larger stimulus sizes.

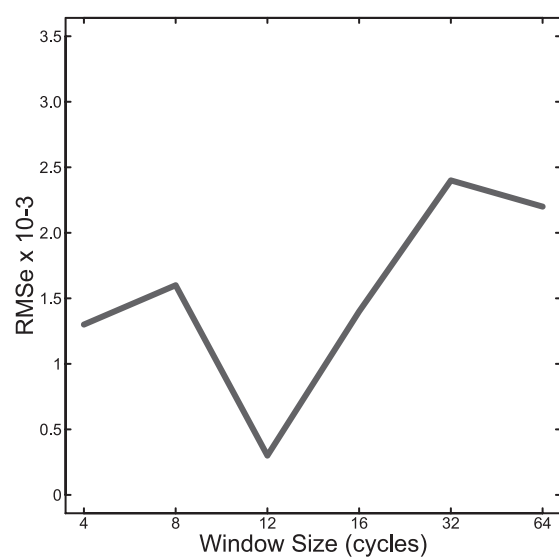


Figure A2. Change in model RMS error according to the radius of the integration aperture of the two-stage model. RMSe decreases as the integration aperture size approaches 12 cycles, and begins in to increase once it exceeds 16 cycles.

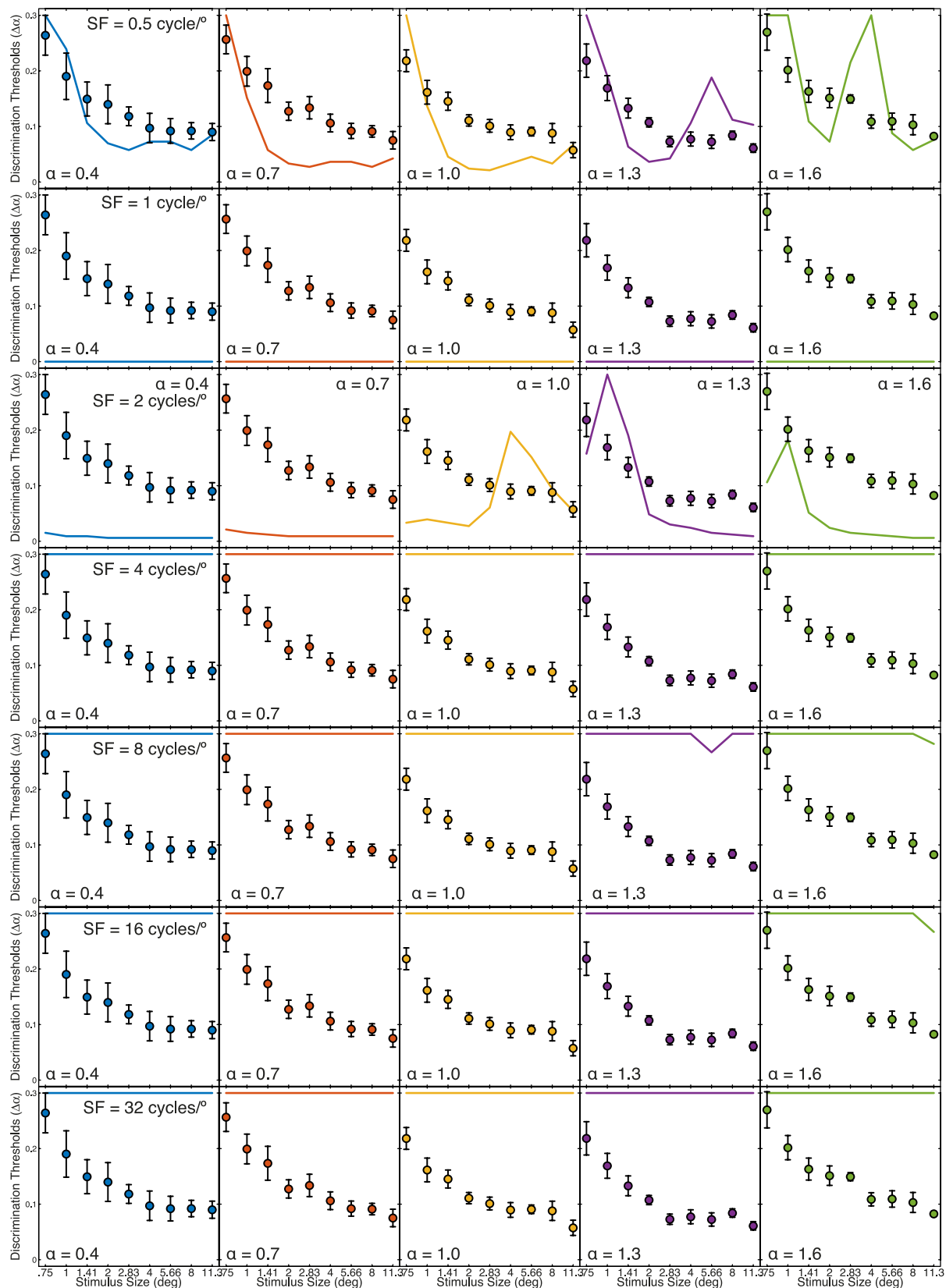


Figure A3. Fits of the single channel model for all spatial frequencies used in this study (0.5–32.0 c/°). The data points mark the discrimination thresholds of observers for a given reference α , separated into subplots (reference α is indicated in the bottom left of the subplot). The single channel model output are shown as lines. Note that for higher spatial frequencies the model was incapable of fitting observer data and thus outputted the highest α discrimination thresholds possible ($\Delta\alpha = 0.3$)