Sensitivity to gaze-contingent contrast increments in naturalistic movies: An exploratory report and model comparison

Thomas S. A. Wallis*

Schepens Eye Research Institute, Harvard Medical School, Boston, MA, USA
School of Psychology, The University of Western Australia, Perth, Western Australia, Australia
Present address: Department of Computer Science and Werner Reichardt Centre for Integrative Neuroscience, University of Tübingen, Tübingen, Germany

Michael Dorr*

Schepens Eye Research Institute, Harvard Medical School, Boston, MA, USA
Present address: Institute for Human-Machine Communication, Technische Universität München, Germany

Peter J. Bex

Schepens Eye Research Institute, Harvard Medical School, Boston, MA, USA
Present address: Department of Psychology, Northeastern University, Boston, MA, USA

Sensitivity to luminance contrast is a prerequisite for all but the simplest visual systems. To examine contrast increment detection performance in a way that approximates the natural environmental input of the human visual system, we presented contrast increments gaze-contingently within naturalistic video freely viewed by observers. A band-limited contrast increment was applied to a local region of the video relative to the observer's current gaze point, and the observer made a forced-choice response to the location of the target (≈25,000 trials across five observers). We present exploratory analyses showing that performance improved as a function of the magnitude of the increment and depended on the direction of eye movements relative to the target location, the timing of eye movements relative to target presentation, and the spatiotemporal image structure at the target location. Contrast discrimination performance can be modeled by assuming that the underlying contrast response is an accelerating nonlinearity (arising from a nonlinear transducer or gain control). We implemented one such model and examined the posterior over model parameters, estimated using Markov-chain Monte Carlo methods. The parameters were poorly constrained by our data; parameters constrained using strong priors taken from previous research showed poor cross-validated prediction performance. Atheoretical logistic regression models were better constrained and provided similar prediction performance to the nonlinear transducer model. Finally, we explored the properties of an extended logistic regression that incorporates both eye movement and image content features. Models of contrast transduction may be better constrained by incorporating data from both artificial and natural contrast perception settings.

Introduction

The visual system's encoding of image contrast has been a central focus of vision science for the past 80 years, and much is known about the topic both psychophysically and electrophysiologically. The visual system is differentially responsive over a range of spatial and temporal frequencies, giving rise to the spatiotemporal contrast sensitivity function (Kelly,

Citation: Wallis, T. S. A., Dorr, M., & Bex, P. J. (2015). Sensitivity to gaze-contingent contrast increments in naturalistic movies: An exploratory report and model comparison. *Journal of Vision*, *15*(8):3, 1–33, doi:10.1167/15.8.3.

doi: 10.1167/15.8.3 Received September 10, 2014; published June 9, 2015 ISSN 1534-7362

1984; Watson & Ahumada, 2005). This sensitivity profile is supported by the combined activity of channels (Campbell & Robson, 1968; Cannon & Fullenkamp, 1991; Graham & Nachmias, 1971; Graham, Robson, & Nachmias, 1978; Haun & Essock, 2010; Meese & Georgeson, 2005; Watson & Solomon, 1997) or neurons (Blakemore & Campbell, 1969; Goris, Putzeys, Wagemans, & Wichmann, 2013; Goris, Wichmann, & Henning, 2009; Kwon, Legge, Fang, Cheong, & He, 2008; Lennie & Movshon, 2005; Ringach, Hawken, & Shapley, 1997; van Hateren & Ruderman, 1998) that respond to different spatial and temporal frequencies. As technologies for stimulus presentation and data analysis have improved, investigators have increasingly used more naturalistic stimuli (Alam, Vilankar, Field, & Chandler, 2014; David, Vinje, & Gallant, 2004; Freeman & Simoncelli, 2011; Geisler, 2008; Geisler, Najemnik, & Ing, 2009; Geisler & Perry, 2009; Haun & Peli, 2013; A. B. Lee, Mumford, & Huang, 2001; Mante, Frazor, Bonin, Geisler, & Carandini, 2005; Peli, 1990; Ringach et al., 2002; Ruiz & Paradiso, 2012; Wallis & Bex, 2012; Wang, Freeman, Merriam, Hasson, & Heeger, 2012) in testing the visual mechanisms elucidated with simplified stimuli in more ecologically valid settings.

There is a great deal of evidence that the visual system's response to contrast differs under more naturalistic conditions compared to simple isolated grating stimuli. In any given natural image, the broad distribution of contrast at different spatial scales (Balboa & Grzywacz, 2003; Frazor & Geisler, 2006) and orientations means that a population of neurons will each be responding at a different level along their contrast response function (Bex, Mareschal, & Dakin, 2007; Goris et al., 2009; Goris et al., 2013). It has been known for some time that these responses are divisively normalized by the activity of nearby channels or neurons, a process called contrast gain control (Foley, 1994; Geisler & Albrecht, 1992; Heeger, 1992; Morrone, Burr, & Maffei, 1982). Here, "nearby" refers to neighboring channels in space, frequency, and orientation.

Recent investigations continue to refine this model. For example, Haun and Peli (2013) found that to account for perceived contrast in broadband images, lower spatial frequencies needed greater weight in the gain pool than high (see also Hansen & Hess, 2012; Haun & Essock, 2010), consistent with the approximately $1/f$ amplitude spectrum slope in natural images. Bex et al. (2007) found that the visual system's contrast response at a given spatial scale is moderated by spectral power at remote spatial scales (cross-scale gain control) in broadband natural scenes, and that the inhibitory influence of the gain pool critically depends on spatial alignment over frequencies (phase coherence). Bex, Solomon, and Dakin (2009) showed that when adapted to natural viewing conditions, sensitivity to contrast at low spatial frequencies is lower than when adapted to a homogenous field. This suggests that simple stimuli interleaved with blank screens overestimate contrast sensitivity for natural conditions. Furthermore, that study found that detection thresholds were relatively uncorrelated with local root-mean-square contrast; rather, higher local edge density was associated with threshold increases. Sinz and Bethge (2013) showed that including a temporal adaptation component, by which the contrast response normalization depended on the recent ambient contrast level, provided a more efficient code for the contrasts in natural images (in the redundancy reduction sense). They tested this model by simulating eye movements on a natural image database, producing a distribution of both similar ambient contrasts (caused by microsaccades) and large changes in ambient contrast (caused by saccades). Finally, Alam et al. (2014) recently showed that a contrast gain control model similar to that of Watson and Solomon (1997) predicted detection thresholds for a vertical log-Gabor target embedded in a natural-image patch remarkably well. Interestingly, the model's predictions were poor in image regions containing identifiable objects (including faces and body parts), suggesting the involvement of additional processing relying on feedback from recognition mechanisms. There is therefore scope for continued improvement of contrast-encoding models by the use of more natural stimuli and tasks.

Experiments in contrast discrimination almost always use static images and have observers maintain stationary fixation. In the real world, we make approximately three fast eye movements per second (e.g., Dorr, Martinetz, Gegenfurtner, & Barth, 2010; Hering, 1879), and objects in the scene move and are often tracked (Wang et al., 2012). To our knowledge, contrast discrimination has never been studied under these conditions, yet they may have large influence over key mechanisms. We examined human contrast discrimination in natural-image sequences (movies) under conditions of free viewing by using a recently developed real-time gaze-contingent display system (Dorr & Bex, 2013) to present image modifications that are contingent on the observer's current gaze position. The movie played without interruption both during and between trials. While not a direct analogue of visual behavior in the real world, this situation is a more realistic way to investigate sensitivity to contrast changes.

Classical investigations of contrast discrimination typically ask observers to discriminate one interval containing a pedestal contrast level from a second interval containing the pedestal plus an increment (Legge & Foley, 1980; Nachmias & Sansbury, 1974). In the simplest cases, the pedestal and the increment are both sinusoidal gratings differing only in amplitude,

and therefore the task is to indicate which grating has higher contrast. Typically, a set of pedestal contrasts would be selected to span the contrast range of interest (including zero contrast, to measure detection), and increment contrasts would be determined either by a method of constant stimuli or using an adaptive procedure to find the threshold increment at each pedestal.

In the present experiment, we incremented the contrast in one spatial band in a local region of the natural-image sequence by a multiplicative factor (see Figure 1). We therefore rely on the variability of local image contrast (see, e.g., Bex et al., 2009, figure 6) in our stimuli to generate a distribution of pedestal contrast samples, which are jointly determined by the content of the video sequence on each frame and the observer's gaze position in the frame. Thus, our pedestal levels span a range of contrasts and take on a large number of unique values. By multiplying the contrast in the unmodified video, we avoid changing the structure of the image at that location, as occurs in using alternative stimuli such as phase scrambling (Bex & Makous, 2002; Oppenheim & Lim, 1981; Sadr & Sinha, 2004; Thomson, 1999; Vogels, 1999; Wichmann, Braun, & Gegenfurtner, 2006) and spatial distortions (Bex, 2010; Rovamo, Mäkelä, Näsänen, & Whitaker, 1997), or imposing new structure such as filtered noise, Gabors, or wavelets (e.g., Alam et al., 2014; Bex et al., 2009; Caelli & Moraglia, 1986; Chandler, Gaubatz, & Hemami, 2009; Eckstein, Ahumada, & Watson, 1997). This method ensured that we measured sensitivity to structure that was as naturalistic as possible.

The results of this study are described in three parts. Part I presents the experimental method and an exploratory data analysis. Part II uses a Bayesian model-fitting framework to examine a commonly used nonlinear transducer model of contrast discrimination; we compare this with an atheoretical generalized linear model (GLM; logistic regression). Part III examines the results of fitting an expanded GLM to the data, combining experimental variables with eye movement and image parameters. This section is intended to provide a starting point for future explorations of this data set, which we have made publicly available.

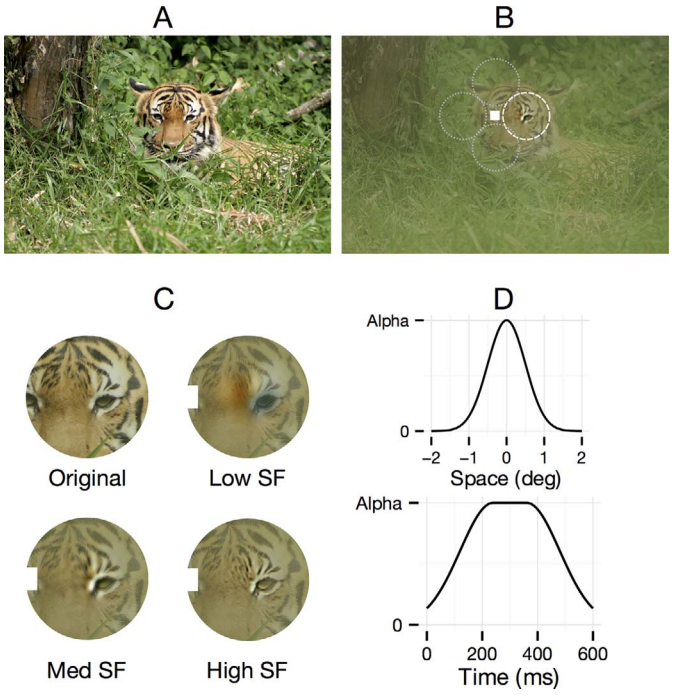

Figure 1. Experimental procedure and stimuli. (A) An image similar in content to the videos used in the experiment. Copyright by Flickr user Phalinn Ooi (2010), released and used here under a Creative Commons 2.0 license. (B) A depiction of our experimental modifications. The global contrast of the image was first reduced (exaggerated here for demonstration). Fixation for this demonstration is at the location of the white square. The contrast increment could be presented above, right of, below, or left of fixation (here right, white dashed circle; other possible target locations shown as gray dashed circles), and the subject responded to the location of the target relative to the fovea (four-alternative forced choice). Here the contrast of the image is incremented at a medium spatial scale. In the experiment, neither the fixation spot nor the dashed circles were presented. (C) Depictions of contrast increments at three spatial scales—low (coarsest), medium, and high (finest)—along with the unmodified patch. The contrast increment was spatiotemporally blended with the movie so no abrupt edges were created. (D) The blending functions used. Alpha represents the multiplication factor that was experimentally varied.

General methods

Observers

The data set for the present analysis consisted of five observers (one woman, four men; aged 20–45 years): two of the authors plus three experienced psychophysical observers. Three additional observers began the experiment but performed relatively few trials due to time constraints; their data have been excluded here in order to reduce the time required to fit the models detailed later. Observers wore optical correction if required. All procedures were approved by the Institutional Review Board of the Schepens Eye Research Institute and adhered to the Declaration of Helsinki. Observers completed a variable number of trials (shown in Table 1).

	0.375–0.75	0.75–1.5	1.5–3	3–6	6–12	12–24	Sum
S1	1,736	1,738	4,313	1,727	1,734	1,725	12,973
S2	717	709	2,298	706	725	722	5,877
S3	436	428	435	437	430	434	2,600
S4	239	227	233	235	240	231	1,405
S5	453	451	456	444	455	421	2,680
Sum	3,581	3,553	7,735	3,549	3,584	3,533	25,535

Table 1. Number of trials for all subjects at each target spatial frequency band (in cycles per degree).

Stimuli and procedure

Subjects were seated 75 cm from a ViewSonic 3D VX2265wm TFT screen (1680 × 1050 pixels; 120-Hz refresh rate) with their heads stabilized by a chin rest. At a size of 47.5 × 29.7 cm, the visual angle subtended by the screen was about 35° × 25°. Blu-ray video stimuli (1920 × 1080 pixels, 23.98 frames/s) were cropped to fit the native display resolution, resulting in a Nyquist frequency of 24 c/° (cpd). Rather than explicitly linearizing the monitor luminance via a lookup-table correction, here we set the monitor gamma to a value of 2.2. The professional video material we used as stimuli has been engineered to be viewed at this gamma correction (Poynton, 2003). Since we do not know the gamma of the camera used to record the footage, or the correspondence between the video intensity values and physical scene luminances (and furthermore, these likely change between scenes), we argue that it does not make sense to linearize the monitor in our case. In the worst case, not doing so may add a small luminance pedestal to our contrast stimuli—equivalent to adding contrast at a lower spatial scale than our intended target. Since luminance and contrast are decoupled both in natural scenes and in their processing by the early visual system (Mante et al., 2005), the scenes vary widely in both luminance and contrast, and the size of the luminance increment is small compared to the contrast increments at high multiplication factors, we do not see that this decision will greatly affect our results.

Eye movements were recorded monocularly using an EyeLink 1000 (SR Research, Ottawa, Canada) eye tracker. Before every experiment session, a 13-point calibration procedure was performed, followed by a 13-point validation scheme. If necessary, these steps were repeated until calibration was accepted by the manufacturer's software.

Each experimental session consisted of a single continuous movie clip. All (non-DC) spatial frequency bands of this image sequence were multiplied by 0.75 to achieve a global image contrast of 75%. The unmodified video was shown for a period of 5–7 s at the beginning of a session after calibrating the eye tracker, in order to set the observers' adaptive state (Bex et al., 2009). A local contrast increment at one spatial frequency band was introduced in one of four randomly-chosen locations centered 2° away from the current gaze position along the cardinal axes (see Figure 1). The contrast increment was smoothly transitioned with the underlying movie by masking the target within a spatiotemporal Gaussian envelope. The spatial support was 2° × 2° with a spatial standard deviation of 0.5°; the temporal support was 600 ms, with the central 120 ms at maximum amplitude and the first and last 240 ms containing a one-sided Gaussian ramp with temporal standard deviation of 120 ms. The amplitude of the contrast increment was determined via a method of constant stimuli, with the band energy of the image at the target patch multiplied by a value between 1.5 and 6.5. The base contrast of 75% ensured that contrast saturation was rare (fewer than 0.5% of pixels saturated).

After the offset of the target, a blue cross (1° × 1°) was presented to the observer's current gaze point to indicate that a trial had just been presented. No cue was presented coincident with target onset. The cross remained on screen, moving with the observer's gaze, until the observer made a button response to indicate the location of the target (four-alternative forced choice). During this period the movie continued to play. After a response, the color of the fixation cross changed to provide feedback, remaining on screen for 400 ms before disappearing. The next trial was initiated with a random delay (1.5–2.5 s) after a response to the previous trial.

Experimental sessions consisted of around 250–300 trials, lasting typically 10–20 min. Movie stimuli were taken from eight episodes (each ≈1 hr) of a popular nature documentary. Sessions were terminated early if an episode reached its end or if the observer noticed that calibration had shifted (by comparing their gaze position to the fixation-cross location). The following session was resumed at the previous movie position.

The position of the local contrast modification on the screen was updated with low system latency to ensure gaze-contingent display to within hardware limits. The image sequence was decomposed online at 24 frames/s (fps) into a Laplacian pyramid with seven scales (Adelson & Burt, 1981). To this end, the image sequence was first decomposed into a Gaussian pyramid by iteratively filtering with a five-tap binomial filter and subsampling by a factor of 2; adjacent scales

were then subtracted (after upsampling the lower scale by a factor of 2) from each other to yield an efficient band-pass representation of spatial frequency, with individual bands ranging from 0.375 to 24 cpd in log steps. Since the lowest level is the DC residual, this yields six band-pass spatial frequency bands. Local modification of one of these pyramid levels before image reconstruction leads to a localized, band-specific change in the reconstructed image. To ensure that image processing was completed within one video frame, we implemented it directly on a high-end graphics processing unit (NVIDIA Tesla C2070) using the C for Graphics shader language (Mark, Glanville, Akeley, & Kilgard, 2003). Pyramid decomposition took 3 ms and local target band modification and image reconstruction took 1.5 ms. The output latency of the monitor, measured with a photodiode, was about 4 ms. The manufacturer-specified latency of the eye-tracking system was 1.8 ms. Thus the overall system latency between an eye movement and its effect on the screen was 18.6–26.9 ms at the center of the screen, depending on eye movement onset relative to the screen refresh cycle. This latency was accounted for in all analyses (time points refer to physical changes on the screen, not the issue time of drawing commands in the computer, which are often confounded in the literature).

Bayesian modeling framework

In Parts II and III of our results, we examine the fits of several models by estimating the distribution of posterior probabilities over the model parameter space—i.e., $p(\theta|D)$, where θ are the parameters in the model and D is the data. The posterior distributions also depend on the prior of the parameters $p(\theta)$ and more generally the selected model architecture. Compared to maximum-likelihood estimation, assessing quantitative models in a Bayesian framework holds a number of advantages, several of which we discuss briefly here. First, estimating the parameters of a model from data using maximum-likelihood methods produces a point estimate of the model parameters, which can be complemented with confidence regions and correlations based on distributional assumptions or resampling procedures. In contrast, the posterior distribution over model parameters provides rich information about the structure of the model and how the parameters relate to one another. Second, it allows uncertainty in the parameters to be preserved in all inferences by using multilevel (hierarchical) models, which are easier to estimate within a Bayesian framework than using maximum likelihood. Third, approximating the posterior distribution using sampling methods (see later) is an extremely general approach that offers great flexibility in the types of models that can be estimated efficiently. We include further discussion of our model fitting approach later in Model estimation and comparison.

Since the posterior distributions for these models are not derivable analytically, we approximate them numerically using Markov-chain Monte Carlo (MCMC) methods. MCMC methods are algorithms that, after they have converged to an equilibrium distribution, generate samples from probability distributions (see Kruschke, 2011, for a useful introduction). That is, the probability of the algorithm's returning a sample at some point in the space is proportional to that point's probability under the target distribution. Highly probable points will be sampled more often. In this case, the probability distribution we are sampling is the posterior distribution over the model parameters. Properties of the posterior distribution, such as marginal means and variances, can be estimated by taking (for example) the mean of the samples from the MCMC algorithm.

Here we use the Stan modeling language (Stan Development Team, 2014) via the rstan package in R (R Core Development Team, 2013), which implements a variant of Hamiltonian Monte Carlo called the No-U-Turn Sampler (Hoffman & Gelman, 2014). All models were fitted using Stan version 2.2.0. Unless otherwise specified, posterior estimates are based on four independent chains using the No-U-Turn Sampler algorithm (Hoffman & Gelman, 2014) and all results reported in this article are based on 5,000 saved samples. Chain convergence was assessed by the $\hat{R}$ value (the ratio of between- to within-chain variance; see Gelman & Rubin, 1992) as well as by inspection of graphical estimates of convergence (Xavier Fernández i Marín, 2014). More details of the model specifications, including priors and sampling parameters, are provided in the Appendix.

Results Part I: Exploratory data analysis

Here we offer exploratory comparisons of performance across several variables of interest, to provide the reader with an overview of several interesting patterns present in the data set. To relate variables of interest, we fitted univariate smoothing splines to the data within a generalized additive model (GAM) framework provided by the mgcv package (Wood, 2011).

Contrast statistics of the stimuli

With pixel intensities of the original image on a [0, 1] scale, coefficients on the pyramid scales are typically small values clustered around 0 due to decorrelation

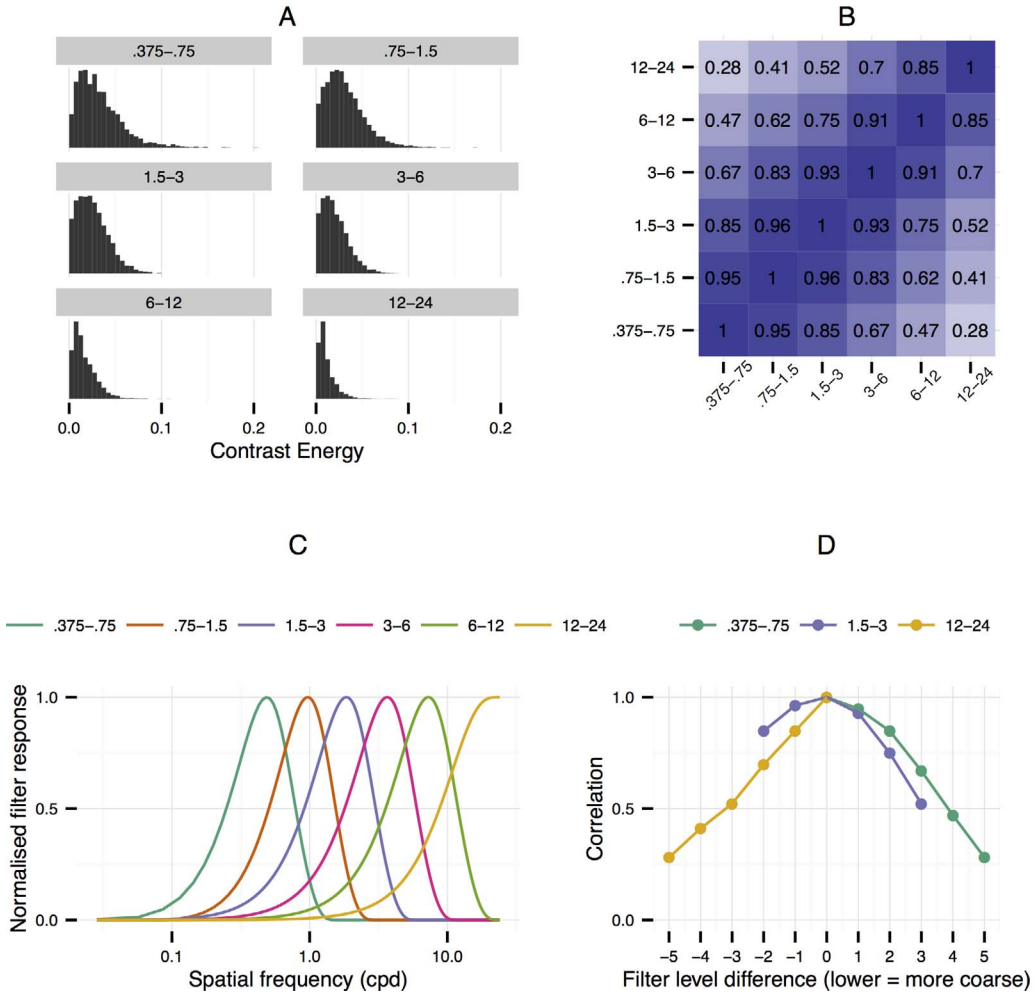

Figure 2. Contrast energy statistics and spatial filtering. (A) Histograms of contrast energy distributions observed in the data set for six spatial frequency bands (each band extent labeled on subplots in cycles per degree). (B) Correlations (Spearman's ρ) between the contrast energy in each spatial frequency band at the target location. (C) Normalized filter responses as a function of spatial frequency. (D) Correlation profiles of three example filters, replotted from (B). The x-axis shows the difference in filter level (lower corresponds to coarser neighbors), and the y-axis shows the rank-order correlation between the example filter (color) and its neighbor.

(Adelson & Burt, 1981). The standard deviation of these coefficients in one frequency band over the target patch ($2° \times 2°$, computed on the unmodified video signal) yields the contrast energy for that band;[1] these distributions and the correlations between them are shown in Figure 2. We consider the relationship between pixel intensity and performance in the Appendix.

It was rare for contrasts to exceed a contrast energy of 0.1, and instances of exactly zero contrast occurred infrequently (21 trials). The bulk of the distribution of pedestal contrast energies in the data set therefore fell between 0.01 and 0.032 (Figure 2A).

The contrast energies in natural signals tend to be correlated across spatial scales (Balboa & Grzywacz, 2000; Field, 1987; Mante et al., 2005; Simoncelli, 1997; Zetzsche, Barth, & Wegmann, 1993). The correlations (Spearman's rank-order coefficient) between the band energies in the six spatial frequency bands observed in our stimuli are given in Figure 2B. Neighboring bands are highly correlated with one another, and this relationship gets weaker with distance (falloff in correlation with filter distance is shown in Figure 2D). The high correlation between nearby bands is not solely due to correlations in natural-scene statistics in our case, however: There is substantial overlap in the frequency selectivity of the pyramid bands (Figure 2C), which will also contribute to the correlations between them.

Performance as a function of contrast manipulation

Targets were created by multiplying the band-limited contrast energy by a factor. Can observers do the task

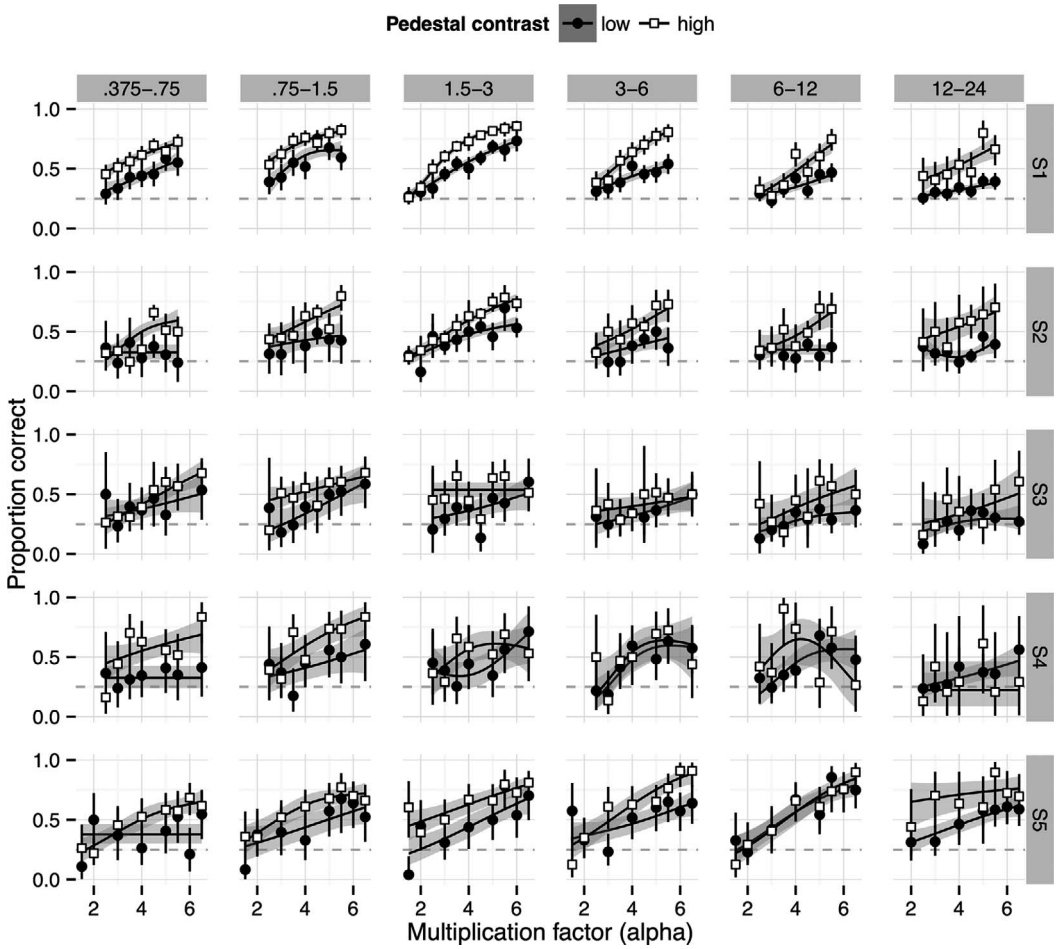

Figure 3. Performance as a function of the multiplication factor applied to the band energy in the target location, for the five observers (rows) at six target spatial frequencies (columns). Target band energy has been binned into two categories for visualisation by median split: low (0–0.019) and high (0.019–0.2) contrast. Points show the expected value for a beta distribution after rule-of-succession correction, and error bars show 95% beta distribution confidence intervals. Curves and shaded regions show the fits and confidence regions of smoothing splines (cubic spline with three knots, GAM with a logistic link function). Dashed horizontal lines show chance performance (0.25).

at all? And if so, what range of performance do they exhibit? Figure 3 shows performance as a function of the multiplication factor, for each observer at each target spatial frequency. In general, performance monotonically increases as a function of multiplication factor, as might be expected. However, performance does not reliably increase until observers are responding perfectly, instead reaching an asymptote at approximately 85% correct for the observer with the most data (S1). The task is inherently difficult, likely due to the spatial and temporal uncertainty of target presentation coupled with the additional uncertainty created by eye movements and the requirement to respond in retinal coordinates (see also Dorr & Bex, 2013). Our conclusions must be considered with the caveat that we have not captured a full range of performance.

Observers also appear to differ on their ability to do the task. For high pedestal contrasts, observers S3 and S4 show less improvement as the multiplication factor increases. Since these observers completed fewer trials than the other observers (see Table 1), it is plausible that this reflects measurement noise or possibly different task strategies.

Another feature of the data noticeable in Figure 3 is that performance differs as a function of the target spatial frequency. Specifically, most observers seem able to perform the task when the target is at a medium spatial scale (0.75–6 cpd), but the slope of performance as a function of multiplication factor decreases for most observers for the higher target spatial frequencies (6–24 cpd). This general pattern of peak sensitivity to contrast is consistent with the typically reported peak of the contrast sensitivity function (1–3 cpd; Campbell & Robson, 1968).

Finally, it can be seen that performance for bins of high pedestal contrast is generally better than for low

pedestal contrasts. We discuss how this result is consistent with contrast masking in the Appendix.

Eye movements

How did observers move their eyes while freely watching naturalistic movies in this task? We collected many hours of eye-movement data that are relevant to television watching, if not classically natural behavior. Overall, eye-movement directions were generally distributed on the cardinal axes (see Figure 4A). Relative to saccades initiated prior to target onset (previous saccades), saccades initiated after the target onset (next saccades) were biased in the direction of the target (more frequent saccades to the "north" in Figure 4A). The amplitude of previous and next saccades did not differ to an appreciable degree (Figure 4B).

A coarse measure of how eye-movement behavior is related to performance is to examine the cumulative distance that the eyes moved over the course of the trial (from target onset to offset). When this value is low, the experimental conditions are a closer approximation to those observed in typical laboratory experiments on contrast discrimination (i.e., the eyes are steady). Performance fell off approximately linearly with the log of the cumulative eye-movement distance in degrees (Figure 4C). This is expected because eye movements will both smear the retinal image and create additional positional uncertainty regarding the target's location relative to the fovea.

Next, we examine the direction of eye movements relative to the target location. We examined how the direction of the next eye movement (initiated subsequent to target onset) was associated with performance (Figure 4D). Saccades in the direction of the target were associated with higher performance than saccades orthogonal to the target's location. Note that this association is correlational: It is possible that eye movements in the target direction were driven by trials in which the observer detected the target then subsequently planned an eye movement in that direction. Alternatively, it is possible that on trials where the observer happened to plan a saccade in the target direction, the observer was more likely to report the correct location of the target, possibly owing to remapping (Deubel & Schneider, 1996).

Some insight into this distinction is provided by the time course relative to target onset (columns in Figure 4D). Saccades initiated between 0 and 120 ms after target onset occur too soon to be in response to the target appearance. Correspondingly, here there is little evidence of a relationship between direction and performance. Conversely, saccades initiated after 120 ms and of an amplitude that would bring the fovea over the screen location previously occupied by the target (1°–3°) show a clear association between direction and performance. These data suggest that the association between saccade direction and performance is likely driven by the observer's first detecting the target then subsequently planning a saccade in that direction.

Figure 4D shows that the detectability of the target depends on the timing of target presentation relative to saccade occurrences. To gain further insight into this relationship, we plot the relationship between performance and the relative timing of the previous and next saccades for saccades within 2 s of the target onset. We find that 91% of previous saccades offset within 2 s before the target onset and 86% of next saccades began within 2 s of target onset. In addition, we show the relationship for saccades *towards* the target (defined here as ±22.5° from the target direction) separately from *other* directions. Figure 5A shows the time from the offset of the previous saccade until the onset of the target. There is little evidence for any relationship between these variables. Conversely, performance appears to depend quite strongly on how soon after the target onset observers made their next saccade. Figure 5B shows the time from the onset of the target until the onset of the next saccade. For saccades towards the target location, trials on which the saccade was initiated around 300 ms after target onset were associated with better performance. This relationship drops before improving again for trials where the next saccade started 1000 ms after the target onset. This may represent trials in which the eye remained relatively still (see Figure 4C). For saccades in other directions, performance seems to slowly decline as the time to saccade initiation increases, peaking around 600 ms, before reversing to join with the "towards" condition again by 1500 ms. The relative dip in performance around 600 ms for saccades in both directions may reflect the influence of geotopic mislocalization errors on performance (Dorr & Bex, 2013). Note that the timing data in Figure 5B correspond to the same variable used to split the direction plots in Figure 4D (columns).

Image features

In our experiment, the image content at the target location depends on what happens to be present in the movie stimulus at any given time. How does performance depend on the features of the image at the target location and surrounding? As an initial approach to this question we consider the *intrinsic dimensionality* of the video signal: the number of dimensions over which the signal changes in a certain spatial or temporal scale. For example, a zero-dimensional signal is one with no change in any dimension. A one-dimensional change could refer to a stationary edge or a spatially

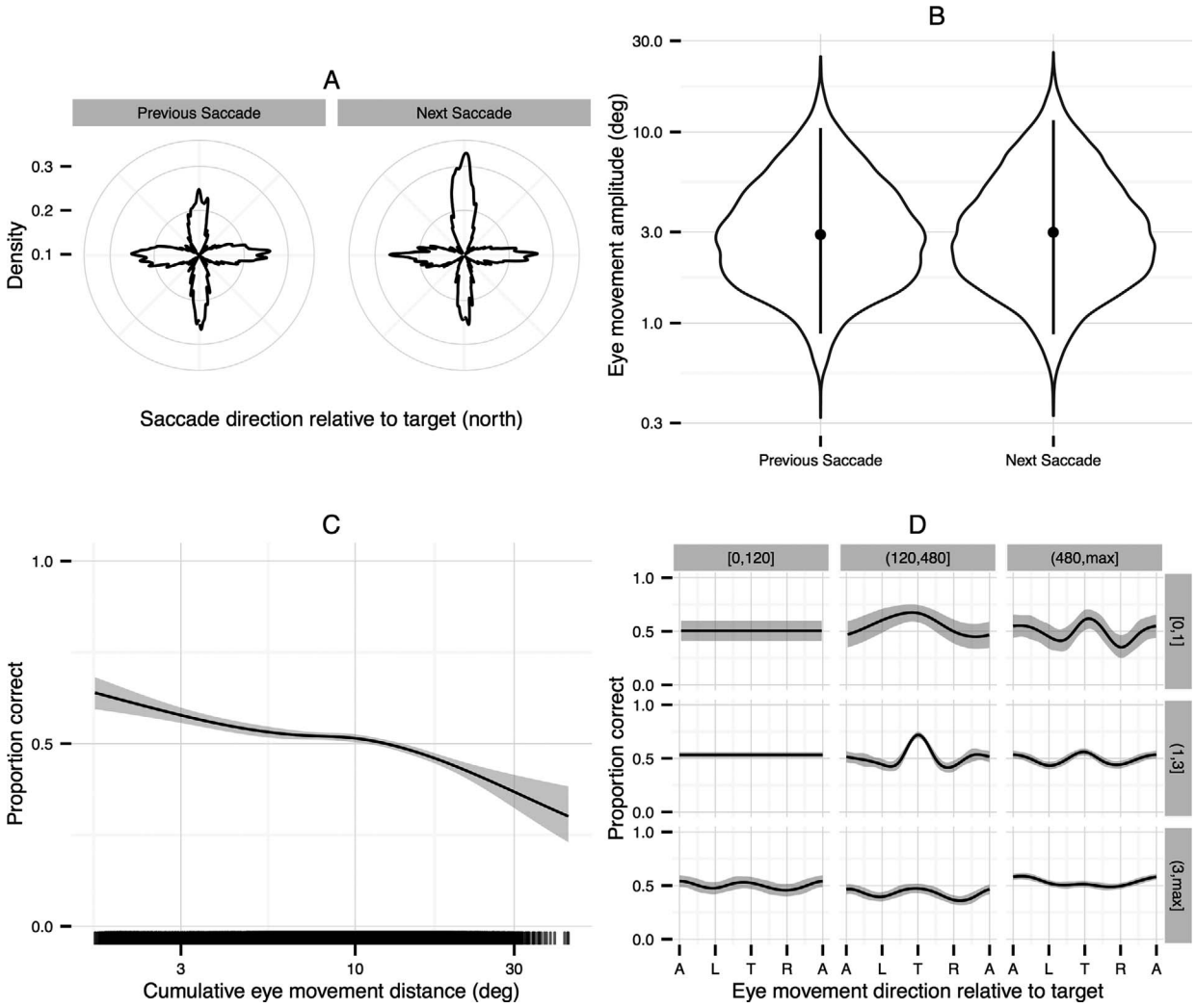

Figure 4. Eye-movement variables and their relationship to task performance. (A) Kernel density estimate of saccade directions, plotted in polar coordinates relative to the target direction (north). Left panel: saccades that were initiated prior to target onset (previous). Right panel: saccades initiated after target onset (next). The kernel is a wrapped normal distribution with a bandwidth of 0.02 radians, estimated using R package Circular (Agostinelli & Lund, 2013). (B) Distributions of saccade amplitudes for the previous and next saccade (note logarithmic y-axis). White blobs are kernel density estimates symmetric about the vertical axis (violin plots). These are superimposed with the mean (points) and 95% highest density interval of the distributions. (C) Performance as a function of the cumulative distance the eyes moved over the course of a trial (note logarithmic x-axis), for all trials with valid eye-movement data (fitted to all data pooled across observers). Solid line shows the fit of a logistic GAM using a cubic spline with five knots; shaded region shows the 95% confidence interval on the fit. Dark dashes on the x-axis show the locations of observed data points. Higher performance was associated with steadier gaze. (D) Performance as a function of the direction of the saccade relative to the target (A = away from, L = left of, R = right of, T = towards the target) initiated after target onset, averaged across observers. The solid lines show fits of a logistic GAM using a cyclic cubic spline with nine knots; shaded region shows the 95% confidence interval on the fit. Panels split the data by time of saccade relative to target onset (in ms, columns) and saccade amplitude (in degrees, rows). The middle panel shows the relationship between performance and saccade direction for saccades initiated between 120 and 480 ms after target onset, that had an amplitude of 1°–3°.

distributed luminance change over time. A two-dimensional change could refer to a corner: The intersection of two edges causes a change in both the x- and y-dimensions. A three-dimensional change refers to a transient corner (one that appears and then disappears) and can only be inferred from a 3-D (spatio-temporal) signal.

The intrinsic dimensionality of a signal can be estimated from the geometric invariants of the structure tensor (Barth & Watson, 2000; Zetzsche & Barth, 1990). In a 3-D signal, if invariant H has a value greater than 0, the signal is *at least* intrinsically one-dimensional; if invariant S is greater than 0, the signal is at least intrinsically two-dimensional; and if invariant K is

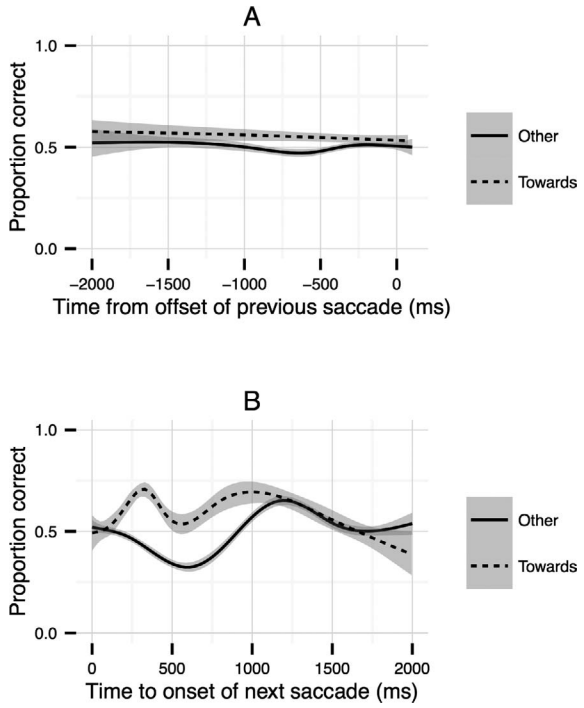

Figure 5. Relationship between performance and saccade timings in relation to target onset (at 0), for saccades towards the target (dashed curve) and in other directions (solid curve). (A) Performance as a function of the time between the offset of the saccade made prior to the target. Solid line shows a logistic GAM fitted to the pooled data across subjects using a cubic spline with eight knots; shaded region shows the 95% confidence interval on the fit. (B) Same as (A) for the time from the onset of the target until the onset of the next saccade. In these plots, a saccade towards the target is one within $\pm 22.5^{\circ}$ of the target's direction relative to the fovea; of the trials analyzed here, this represents 15% and 19% for the previous and next saccades, respectively.

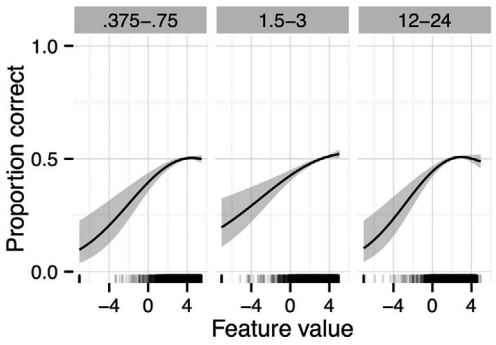

Figure 6. Performance as a function of spatiotemporal feature energy, defined as the (log-scaled) value of geometric invariant *K* computed in three dimensions (i.e., across both space and time), at three spatial scales for a moderate temporal window (160 ms). Higher feature values correspond loosely to more intense transient corners. Loosely, this can be considered a measure of edge density or spatiotemporal structure in the target location. The dashed dark lines above the x-axis show the density of trials: Most trials cluster between log feature values of 0 and 5 (note also the uncertainty on the spline estimates lower than feature values of 0). These spline fits use a lower asymptote of 0 and so do not consider chance performance in the task; the lower bound of these functions is therefore misleading.

greater than 0, the signal is intrinsically three-dimensional. We average responses of these geometric invariants over the target patch; loosely, this can be thought of as an index of feature density (where features are edges in space and time). In order to capture features of different sizes, we computed the invariants on an anisotropic spatiotemporal pyramid with six spatial and three temporal scales (i.e., over the spatiotemporal movie signal). As for the contrast statistics, the invariants were computed on the unmodified video signal within a $2^{\circ} \times 2^{\circ}$ region centered on the target location, linearly averaged over 14 frames of the video signal centered on the target's contrast peak. We additionally computed 2-D invariants (i.e., for static movie frames) and also invariants at the nontarget locations, but we do not present them in this article. They are provided in the data set online.

Based on informally exploring the relationship between these invariants and task performance, we selected the invariant *K* computed over the three-dimensional video signal at three spatial scales and over a moderate temporal window (160 ms). The relationship between (log) feature intensity and performance for these invariants is shown in Figure 6. At all spatial scales, performance increases as the feature value increases above very low values. For invariant *K* at coarse (0.375–0.75 cpd) and fine (12–24 cpd) spatial scales, the relationship between the feature value and performance then plateaus or even begins to decrease over the range of the bulk of the data. For three-dimensional change over a moderate spatial scale (1.5–3 cpd), the relationship between performance and log feature intensity continues to increase approximately linearly over the range of the data.

To what extent do the invariants give us information about the image structure that is not provided by contrast alone? In Figure 7 we show the relationship between contrast energy and spatiotemporal feature energy in the 1.5–3 cpd range (as in Figure 6, middle panel), for trials in which the target was presented at that spatial scale. While the two quantities are moderately correlated (Spearman's rank-order correlation = 0.54), the spatiotemporal feature energy does provide additional information. Intrinsically 3-D structure (i.e., transient corners) can be observed even when the overall contrast is low. Thus the invariants do provide information in addition to simple contrast (which is also correlated with edges in natural scenes).

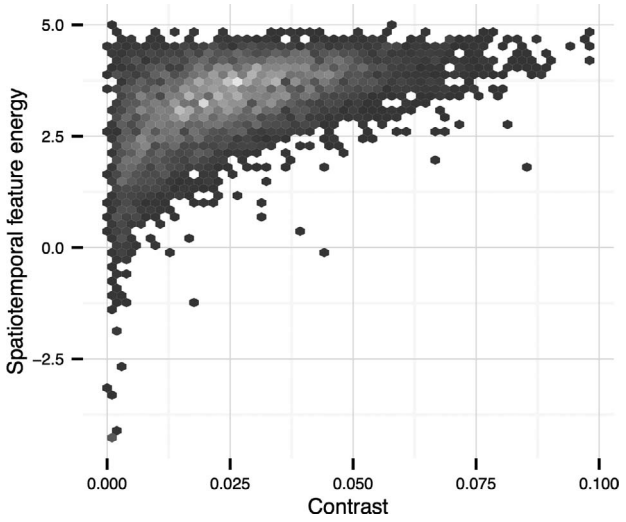

Figure 7. Relationship between contrast energy and spatiotemporal feature energy (invariant *K*, log scaled) at the target location in the 1.5–3 cpd range, on trials where targets were at 1.5–3 cpd. Individual data pairs were aggregated into hexagonal bins to reduce overplotting; lighter bins represent more samples. The Spearman (rank-order) correlation between contrast and invariant *K* is 0.54. Note the semilog axes.

Results Part II: Nonlinear transducer models and GLMs

In the previous section we presented an exploratory data analysis to demonstrate several patterns evident in the data. In this section we ignore the influence of eye movements and the content of the image at the target location (apart from simply the contrast), and consider the data from the perspective of a pure contrast increment detection task. Since Figure 3 revealed that the observers have variable performance and that most cannot reliably develop good performance for targets presented in spatial bands above 6 cpd, in this section we consider data from only the 1.5–3 cpd condition (i.e., 7,735 trials). We describe the properties and prediction performance of two versions of a simple nonlinear transducer model and an atheoretical GLM.

Nonlinear transducer models

Method

The standard framework for considering contrast increment detection is to model the system's internal contrast response with some form of nonlinear transducer model, after Legge and Foley (1980) and Stromeyer and Klein (1974). The visual system is hypothesized to produce an internal response to a given level of contrast. The ability of the observer to discriminate two contrasts is determined by the difference between the responses produced by each contrast. Contrast detection is a special case where one of the contrasts is zero. This basic framework has been very successful in modeling data from numerous detection and near-threshold discrimination experiments. Here we apply two simple variants of this model class to the contrast increment detection data from our experiment.

The task for our observers is to discriminate which spatial location contains an (unnatural) increment in contrast. The naturally occurring (unmodified) contrast at each spatial location can be considered the pedestal contrast: the base level from which the observer must discriminate the increment (see Discussion for further elaboration of this assumption). The contrast after modification is the "pedestal plus increment" contrast. Performance is determined via the hypothetical internal response difference between the two:

$$\Delta_R = R_{\text{ped+inc}} - R_{\text{ped}}. \quad (1)$$

The response difference is then related to the probability of a correct decision via the signal detection theory framework. The probability of a correct response in a forced-choice task is given by

$$\text{p(correct)} = \int_{-\infty}^{\infty} dx \phi(x - \Delta_R) \Phi(x)^{m-1}, \quad (2)$$

where ϕ is the unit normal density function, Φ is the cumulative Normal density function, and m is the number of alternatives in the forced-choice task (Hacker & Ratcliff, 1979). To avoid needing to evaluate this integral at every step of the MCMC process, we use a Weibull approximation to this function (see Appendix).

How is the internal response R determined? The contrast response function is a modified Naka–Rushton sigmoid, which has a general form consisting of an excitatory component E and a divisive inhibitory component I:

$$R(c; \text{E}, \text{I}) = \frac{E(c)}{I(c)}. \quad (3)$$

In the nonlinear transducer of Legge and Foley (1980), both the excitation and the inhibition are functions of the contrast in the target channel. We use the following parameterization of this model:

$$R(c; p, q, z, rmax) = rmax \frac{c^{p+q}}{z^p + c^p} \quad (4)$$

where c is the contrast at the target location and spatial band and R is the output response. The function has four parameters; z is the semisaturation point, in effect determining the horizontal position of the curve with respect to contrast. If z is 0, then there is no accelerating portion of the nonlinearity. The argument

rmax is a scaling factor that determines the maximum absolute response. The shape of the curve is determined by p and q (accelerating and decelerating portions). If q is 0, the function is a familiar sigmoidal psychometric function with a 50% point of z and a maximum asymptotic value of *rmax*; p determines the steepness of the function. If q is greater than 0, the response continues to increase as a function of contrast with an exponent of q once contrast is greater than the threshold (z). The four parameters are related to one another in nontrivial ways (shown analytically by Haun, 2009; Yu, Klein, & Levi, 2003).

It is informative to consider the fits of two transducer models with different priors. After Legge and Foley (1980), parameters p and q are often set to be around 2 and 0.4, respectively (e.g., Alam et al., 2014; Haun & Peli, 2013). These values give the dipper shape (see Solomon, 2009, for a broad tutorial) that provides a good fit to the facilitation effect found robustly in classical contrast increment detection tasks using sinusoidal gratings, as well as for broadband noise stimuli (Henning & Wichmann, 2007), $1/f$ noise patterns (Haun & Essock, 2010), and static natural images (Bex et al., 2007). Note that the exponent values depend on task conditions (Haun & Essock, 2010; Kwon et al., 2008; Meese & Holmes, 2002; Wichmann, 1999; Yu et al., 2003). For example, Wichmann (1999) found that the exponents depended strongly on the exposure duration for grating stimuli. We do not consider these effects further here.

In our Bayesian modeling framework, fixing parameters to certain values is equivalent to placing an infinitely tight prior distribution over the parameter—for example, a normal distribution centered on the relevant value with a standard deviation of 0. The first model we consider here (Transducer A) uses relatively weak priors over all four parameters, allowing the data to have a large influence on the posterior. By contrast, the priors in Transducer B are very strong for all but one parameter, severely constraining the posterior as in previous literature. Details of the structure of the priors are provided in the Appendix.

For these two models, we fitted the four parameters from the nonlinear transducer model (Equation 4) to the data from our five observers, estimating each parameter separately for each observer, using the MCMC package Stan (see previous). That is, there is one parameter in the model for each subject i: p_i, q_i, z_i, and $rmax_i$. The outcome of each trial (correct or incorrect) was assumed to be a Bernoulli random variable with probability given by our Weibull link function (Equation A2). The contrast c in Equation 4 used to calculate R_{ped} (the pedestal response) was the band energy at the target location in the target spatial frequency band, and the contrast for $R_{ped+inc}$ was the R_{ped} contrast multiplied by the alpha level from the trial.

Results

The results from this model fitting are summarized in Figure 8. Figure 8A shows the priors and posteriors from the Transducer A model. The violin plots (kernel density estimates symmetric about the vertical axis) reveal strong bimodality in the posterior distributions for many parameters. For example, posterior estimates for parameter q for S1 and S2 (the subjects with the most data) have modes located at approximately 0.3 and 0.5; similarly, the posteriors for *rmax* appear strongly bimodal. In addition, posterior distributions for p are highly skewed, with the bulk of the distribution tending towards 0 but with a large range. These bimodalities are driven by correlations between the parameters of the model, a point that we return to later (Figure 9).

Figure 8B is the same as 8A, but for Transducer B. First consider the prior distributions in both cases: Whereas in Transducer A the priors were relatively broad (for example, the prior on *rmax* was a uniform distribution ranging 0–100), in Transducer B the priors for p, q, and *rmax* are extremely tightly centered over values of 2, 0.4, and 10, respectively. Correspondingly, in this model the posterior distributions for these three parameters remain similar to the priors (i.e., the data do not have great influence on the strong priors). Only the prior on z (uniform 0–1) is the same as in Transducer A. Compared to the posterior for z in Transducer A, the posterior for Transducer B is very tightly constrained; this is driven by the data and the fact that the other parameters are also relatively fixed.

The performance on a single trial is given as a function of the pedestal contrast and the increment contrast (Equation 1). In Figure 8C we show the model predictions (posterior mean) over a surface of pedestal versus increment contrast for each transducer model. The dashed lines in this figure show iso-performance contours across this surface corresponding to d' values of 1, 1.5, and 2.5. These curves are equivalent to threshold-versus-contrast (TvC) functions. For Transducer A it can be seen that the TvC functions rise smoothly as a function of the pedestal contrast. This is a simple masking function: As the pedestal contrast increases, a larger increment is required to reach the same level of performance. In contrast, Transducer B exhibits the characteristic dipper shape, in which performance first *improves* relative to very low pedestal contrasts (i.e., detection) before rising sharply into a masking curve. The TvC functions for Transducer B are shaped this way because our strong priors force them to be. Interestingly, the dipper occurs at a high pedestal contrast relative to the range of the data, and the model over the bulk of the data (indicated by the red and blue density contours) shows a very different shape to Transducer A.

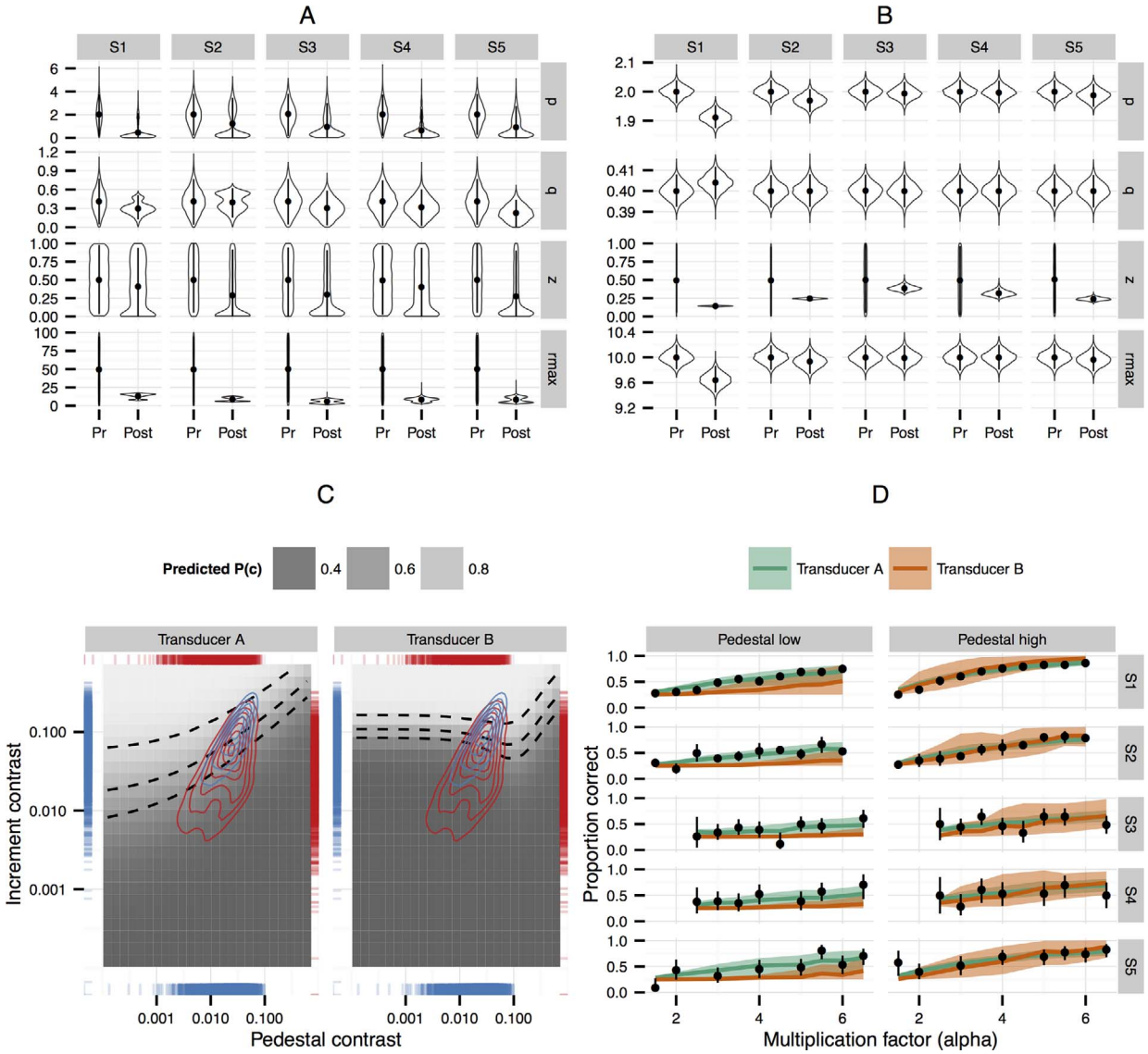

Figure 8. Results for two nonlinear transducer models fitted to the data from the 1.5–3 cpd condition. (A) Prior (Pr) and posterior (Post) samples for each parameter in the Transducer A model, for each subject. Sample distributions are shown as violin plots, which are kernel density estimates symmetric about the vertical axis. Points show the mean of the samples and bars show the 95% highest density intervals. Note the very different y-axis scalings. (B) Same as (A) for the Transducer B model. (C) Predicted proportion correct (posterior mean) as a function of the pedestal contrast and the increment contrast for the Transducer A and B models, subject S1. Predicted proportion correct on the surface is shown in grayscale values, where darker colors represent lower predictions. Dashed curves show iso-performance contours over the surface corresponding to d' values of 1, 1.5, and 2.5 (approximately 55%, 70%, and 91% correct). These are equivalent to threshold-versus-contrast (TvC) functions. The locations of correct (blue) and incorrect (red) trials in this space are shown both as rug plots (dashes near axes) and as two-dimensional density estimates. The axes of these plots have been adjusted to show the main range of the data; several trials at low pedestal contrast are excluded. (D) Performance as a function of the multiplication factor alpha, for high and low pedestal contrasts (data replotted from Figure 3, 1.5–3 cpd condition). Green and orange curves depict the posterior mean of predictions for the Transducer A and B models, and the shaded regions show the 95% highest density intervals of these predictions. See text for details.

Figure 8D shows predictions of the transducer models for the data as a function of the multiplication factor (replotted from Figure 3). The prediction curves were calculated by generating a predicted proportion correct for each trial in the experiment for each saved MCMC sample, and then calculating the mean and 95% highest density interval of these predictions for each subject at each alpha level. The predictions in this plot are not smooth (as in Figure 8C), because the multiplication factor was not explicitly parameterized in the model, so predictions represent an average over different pedestal contrasts. Transducer A provides a

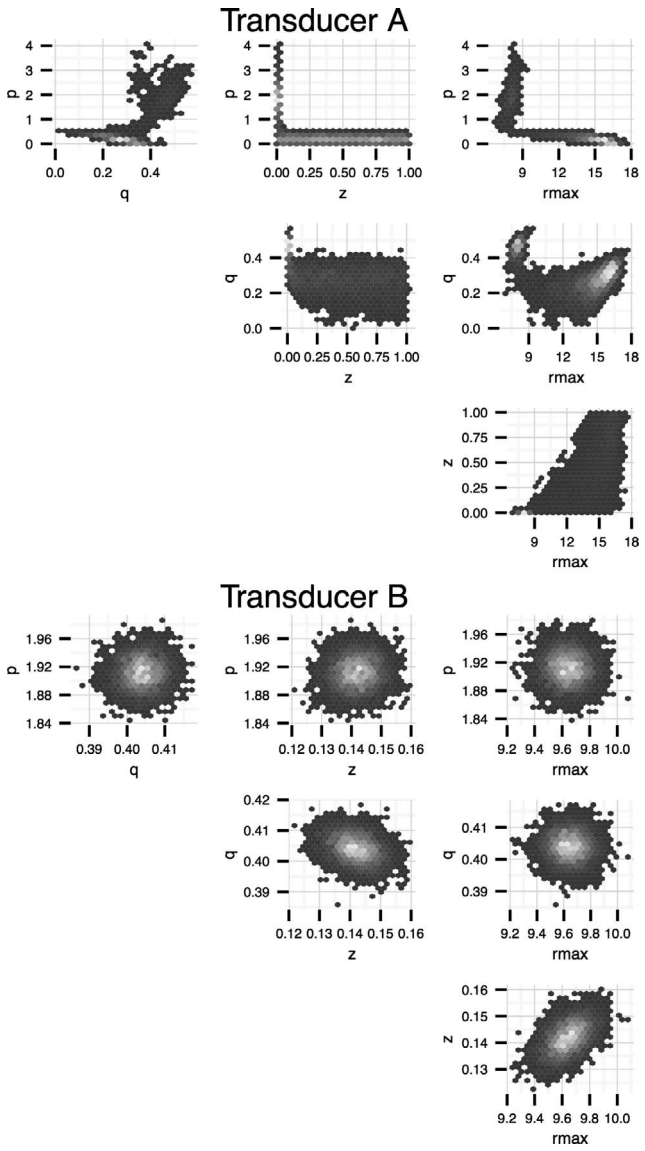

Figure 9. Scatterplot matrices showing the bivariate distributions between each model parameter of the Transducer A (top) and Transducer B (bottom) models, for observer S1. The 5,000 MCMC samples were aggregated into hexagonal bins to reduce overplotting; lighter bins represent more samples and thus a higher posterior density.

reasonable fit to the data. Conversely, Transducer B severely underestimates performance at low pedestal contrasts, and while its mean prediction is similar to that of Transducer A at high pedestal contrasts, the prediction variance is much higher.

As noted earlier, estimating the posterior distribution of a model via MCMC methods provides rich information about the model structure. Here we show various views onto the posterior distribution of the two transducer models and discuss notable features. We do so using a type of scatterplot matrix showing the bivariate distributions of each parameter plotted against each of the other parameters in the model (these are shown in Figure 9). These are the posterior distributions presented in Figure 8A and B replotted to show the relationships between the parameters.

The top panel of Figure 9 shows this data for Transducer A, whose priors were left relatively weak. The posterior over model parameters is poorly behaved, with evidence for bimodality, correlations between parameters, and extremely non-Gaussian shapes. This result is expected from the analytically derived dependencies between the model parameters (Haun, 2009; Yu et al., 2003). It is strong evidence that our data do a poor job of constraining the parameters of this model. This impression was confirmed in several simulation studies (not presented here), in which we found that maximum-likelihood techniques yielded unstable estimates of the model parameters, indicating a flat likelihood surface with multiple local minima.

In contrast, the bottom panel of this figure shows that the posterior distribution for Transducer B appears approximately Gaussian, and the parameters are largely uncorrelated with one another. This is by design, since in this model the prior distributions were strong and the parameters were assumed to be independent (see Appendix). The exception is for the parameters z and *rmax*, which do show evidence of a moderate positive correlation. This is because the z parameter was the only one given a broad prior, and its influence on the contrast response profile is traded off against *rmax*.

Before considering how well the nonlinear transducer models predict the data more formally, we first introduce two versions of a GLM (logistic regression) using the same predictor variables.

Generalized linear models

Method

An atheoretical approach to modeling the data from the current experiment is to apply a GLM in the form of a modified logistic regression. In this framework, the sum of multiplicatively weighted predictor variables is passed through a (logistic) link function that transforms the unbounded weighted sum into the range of [0, 1], which is then the expected value of a Bernoulli trial. Formally, the linear predictor η is given by

$$\eta = \beta_0 x_0 + \beta_1 x_1 + \cdots + \beta_n x_n, \quad (5)$$

where the β values are weights and the x_n values are the individual predictors. This can also be described in matrix notation as $\eta = \beta^{\mathrm{T}}\mathbf{X}$, where rows in $\mathbf{X}$ are trials. The linear predictor is then passed through a modified inverse logistic link function

$$\mathrm{p(correct)} = \gamma + \frac{(1-\gamma)}{1+\exp(-\eta)}, \quad (6)$$

where γ is a lower bound of performance, in this case 0.25. The key difference between the GLM considered here and the earlier nonlinear transducer models is that the calculation of the internal response R involving several exponents and a division (Equation 4) has been replaced with the simpler linear combination of predictor variables.

We first consider a GLM with three predictor variables—the pedestal contrast, the increment contrast, and an intercept—that are estimated for each subject separately (this model is the *single-level GLM*). Any pedestal or increment contrast values of 0 were set to the minimum nonzero value, then the log of the contrast was taken. The design matrix of the model fitted to the trials of each subject i is then given by

$$\eta_i = \beta_{i,0} + \beta_{i,1}\text{ped} + \beta_{i,2}\text{inc}, \quad (7)$$

where ped is a vector of the log pedestal contrast on all the trials from subject i, inc is the log increment contrast on each trial, the first coefficient $\beta_{i,0}$ is the intercept, $\beta_{i,1}$ and $\beta_{i,2}$ are the slopes of ped and inc, respectively, and η_i is a vector of linear predictor values for each trial. We then normalized the design matrix by subtracting the mean and dividing by the standard deviation (i.e., z-scores were computed). Normalizing the predictors makes the model easier to interpret because the intercept represents the level of performance when pedestal and increment contrast were at their mean in the data, and the magnitudes of the coefficients are comparable since they are based on z-scores.[2]

Second, we consider a *multilevel* extension of this model, such that each subject is considered as part of a population and the individual-subject β coefficients are estimated concurrently with the mean and variance of the coefficients at the population level. Multilevel (hierarchical) models have the desirable property of creating *shrinkage* (also called *partial pooling*) between parameter estimates: Each subject's coefficient estimates are influenced by those of the other subjects to the extent suggested by the data, via the population variance term (Gelman, 2006; Gelman & Hill, 2007; Kruschke, 2011). Observer-level parameters with higher variance (greater uncertainty) will be pulled more strongly towards the center of the population distribution than observer-level parameters with low variance. This is a conservative influence on inference, since it can reduce false alarms. The multilevel models considered here are a more general version of mixed models, which have recently been advocated for analyzing psychophysical data (Knoblauch & Maloney, 2012; Moscatelli, Mezzetti, & Lacquaniti, 2012). The multilevel model is somewhat like estimating separate regressions for each subject, then conducting another regression on the individual-subject coefficients to derive population estimates, except that here all these parameters are estimated concurrently. In addition to the individual-subject $\beta_{i,j}$ terms in Equation 7, the multilevel model has a population mean μ_j and standard deviation σ_j for each predictor variable j, whose posterior distributions are also estimated.

For both the single-level and multilevel logistic GLMs, we use weakly informative prior distributions over the parameters. For example, in the single-level model, the prior for each β is a normal distribution with mean 0 and standard deviation 2. Since the predictors are in standard units, this represents a broad a priori range of slope values. The two GLMs are developed in more detail in the Appendix.

Results

The results from fitting the single-level GLM (i.e., independent parameter estimates for each subject) and the multilevel GLM are summarized in Figure 10. This figure is similar in its structure to Figure 8.

Considering Figure 10A and B, it can be seen that relative to the (broad) prior distributions, the posterior distributions are tightly constrained by the data. Furthermore, unlike the posterior distributions for the Transducer A model in Figure 8, these distributions appear well behaved (no evidence of bimodality). In terms of the model coefficients, note that, since this is a linear model, a coefficient of 0 means that the parameter has no relationship to performance (i.e., it is multiplied by 0 and so dropped from the model). The coefficients for pedestal contrast are all negative on average. This indicates that the probability of success on a trial decreases as the pedestal contrast increases, and is indicative of contrast masking (consistent with Weber's/Stevens's law). On the other hand, the coefficients for the increment contrast are all strongly positive, which is no surprise because this is effectively the signal strength of the target: As increment contrast increases, the probability of getting the trial correct also increases. Since the predictors are all normalized to have mean 0, the intercept parameter represents the average level of performance (in the linear predictor) for each subject when pedestal and increment contrast were at their mean.

Figure 10C shows that the iso-performance contours of the models (i.e., the TvC functions) are linear through log-pedestal versus log-increment space. This is a property of the models, since they contain no interaction terms or higher order polynomials that would allow curvature in the model surface.[3] The predictions of these atheoretical models do not make much sense outside the range of the data. For example, they predict that as the pedestal contrast becomes increasingly close to zero, increment contrast also approaches zero. This is not realistic, since we know that the contrast increment detection threshold is

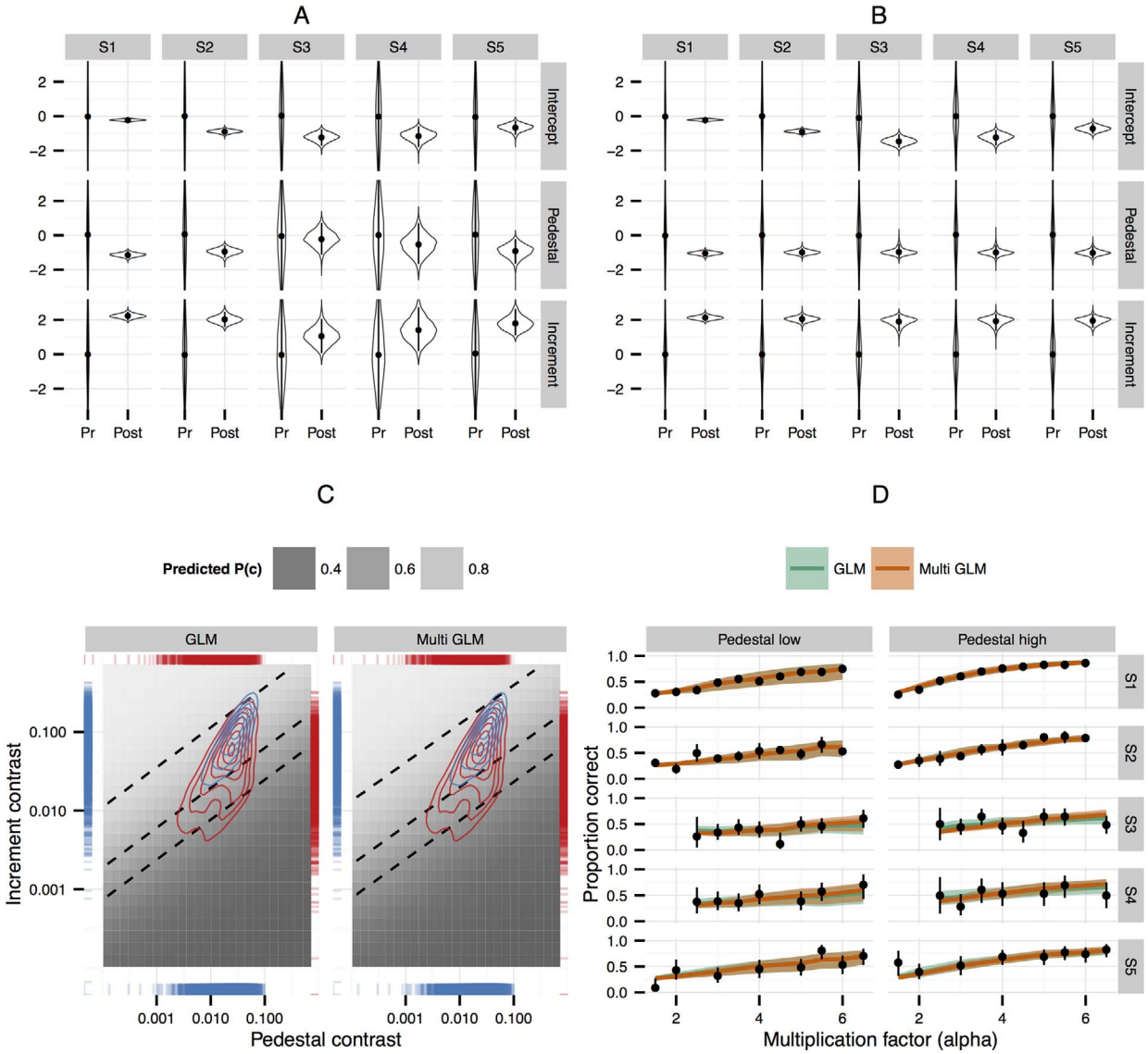

Figure 10. Results for two logistic GLMs fitted to the data from the 1.5–3 cpd condition. Display as in Figure 8; refer there for explanation. (A) Prior and posterior samples for the single-level GLM. Note that the y-axis range has been reduced to show the useful range of the posterior marginal distributions; prior distributions extended from −5 to 5. (B) Same as (A) for the multilevel GLM, in which coefficient estimates from individual subjects influence each other. (C) Predicted proportion correct as a function of the pedestal and increment contrast for the single-level and multilevel models, for subject S1. (D) Predictions as a function of multiplication factor, along with data replotted from Figure 3. See text for details.

appreciably above zero—i.e., it must saturate at some lower bound. Nevertheless, note that within the range of the data of the present experiment (density contours in Figure 10C), the iso-performance contours are similar to those of Transducer A in Figure 8C.

Finally, Figure 10D shows the model predictions for performance as a function of the multiplication factor at high and low pedestal contrasts (binned as in Figure 3). The posterior mean predictions (curves) are quite similar between the two models, but the single-level model shows higher prediction variance. This is because in the multilevel model, all the data are used to inform the fits at the subject level, meaning that the coefficients for subjects with less data are moved towards the population mean.

Figure 11 shows the posterior distributions of the two GLMs as in the nonlinear transducer models before. As for the Transducer B model, the posteriors for both GLMs are quite well behaved, showing approximately Gaussian shapes, albeit with relatively strong correlations between the parameters. The difference between this and Transducer B is that here the prior distributions were relatively weak, whereas there the prior distributions strongly constrained the model.

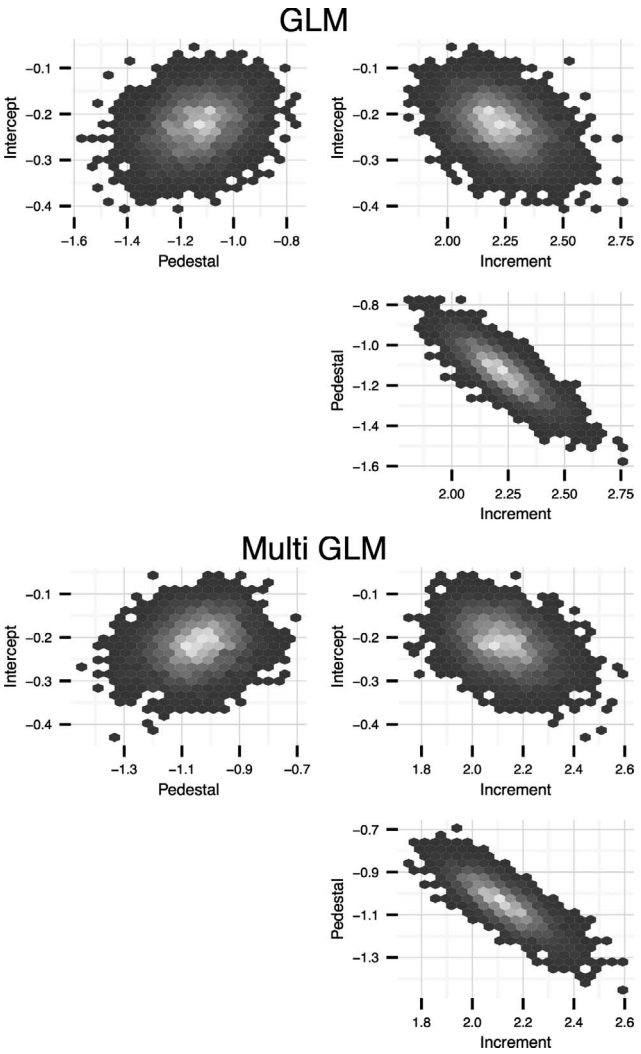

Figure 11. Scatterplot matrices showing the bivariate distributions between each subject-level model parameter, for the single-level (top) and multilevel GLMs (bottom), for observer S1. As in Figure 9, the 5,000 MCMC samples were aggregated into hexagonal bins to reduce overplotting; lighter bins represent more samples and thus a higher posterior density.

Model comparison: Predictive performance

We have described four models fitted to the trials from the 1.5–3 cpd condition: two nonlinear transducer models with different priors, one single-level GLM (fitted to each subject separately), and one multilevel GLM (including partial pooling across subjects by estimating population-level parameters). How well do the models predict unseen data? We fitted the four models in a cross-validation framework. The 7,735 trials in this condition were partitioned into five folds, each model was fitted to all but one fold, the remaining (unfitted) trials were predicted from the model coefficients, and then this procedure was repeated for each fold. We thereby obtained a prediction (expected proportion correct) for each trial where that trial was not included in the model fitting; this gives a more realistic estimate of the out-of-sample predictive performance of the model.[4] The predicted proportion correct was the posterior mean, calculated by generating a prediction for each MCMC sample and then taking the mean over samples for each trial.

We examine two measures of predictive performance. The first is the area under the receiver operating characteristic (ROC) curve (a plot of hits versus false alarms as the threshold for classifying as success or failure is raised from 0 to 1). If the models had no predictive value, the area under the ROC would be at 0.5, whereas perfect prediction would be 1. Values of the area under the ROC were calculated using the ROCR package (Sing, Sander, Beerenwinkel, & Lengauer, 2005). The second is the average log likelihood of the data, measured in bits per trial (i.e., taking the logarithm with base 2). This score is calculated relative to a baseline model in which the mean performance from the training set was used as the prediction for the test set. Each model log likelihood is presented as the difference from this baseline model. A score of 1 bit/trial relative to the baseline model would mean that the model in question predicts the data twice as well as the baseline ($2^1 = 2$). The average log likelihood is a more continuous measure of prediction performance than the area under the ROC, since it measures the distance of the data from the prediction rather than just the sign. Note that the baseline model performs at 0.5 for the area under the ROC curve metric because it provides no information on whether individual trials lie above or below the mean.

The prediction results are shown in Figure 12. To provide some intuition about the uncertainty of the model predictions, we also show bootstrapped 95% confidence intervals on the log likelihoods. The single-level GLM, multilevel GLM, and Transducer A show similar predictive performance. Transducer B, with parameters tightly constrained to values used in previous experiments, shows relatively poor prediction. In terms of log likelihood, it is about the same as the baseline model. The ROC and log likelihood show a similar pattern of results. To verify our MCMC results, we also present the predictions of the single-level GLM fit via maximum-likelihood methods, using a custom logistic link function (equivalent to Equation 6) from the psyphy package (Knoblauch, 2014). This is shown as the dashed horizontal line in Figure 12. It overlaps with the GLM, providing a check that our MCMC methods are working.

Two caveats should be borne in mind when considering these predictive performances. First, our simple bootstrap procedure assumes that all observations are independent, while our data have dependencies both within and across observers. Second, cross

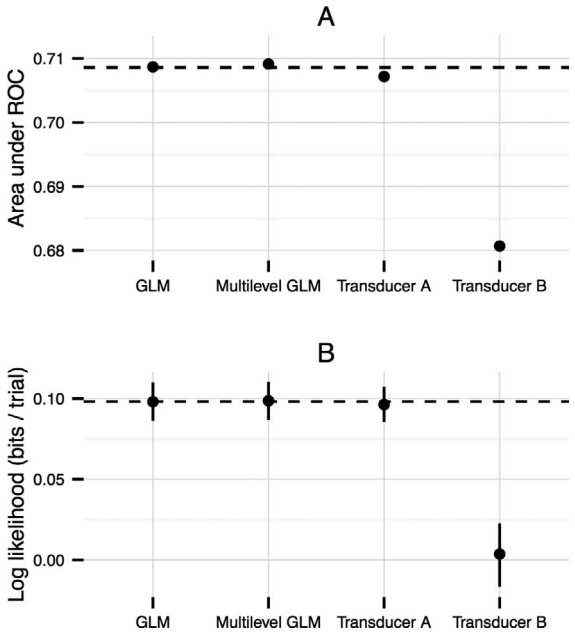

Figure 12. Predictive performance for the four models assessed via fivefold cross validation, for two metrics. (A) The area under the ROC curve (higher is better, chance is 0.5). (B) The average log likelihood in bits per trial, relative to a baseline model in which the mean performance from the training set was used as the prediction. Positive values indicate improvement over the baseline model. Horizontal axis shows the models (single-level GLM, multilevel GLM, and nonlinear Transducer A and B models, respectively). Dashed horizontal line shows the performance of a single-level GLM (equivalent to the "GLM" model) estimated by maximum likelihood. Error bars show 95% confidence intervals derived from 5,000 bootstrap samples.

validation is performed on the individual-trial level rather than between subjects. Consequently, the results in Figure 12 are likely to underestimate both prediction errors and uncertainty, particularly for predicting the behavior of a new subject. Nevertheless, the bootstrapped confidence intervals give a sense of the uncertainty in these estimates, and additionally show that the average prediction performance of all test models are well above baseline except nonlinear Transducer B.

Results Part III: GLM incorporating other predictors

We have shown that a GLM (logistic regression) has similar predictive performance to the more mechanistic nonlinear transducer model and is arguably easier to interpret, because its parameters are less dependent and more constrained by our data. In this section we seek to include both the contrast information and some non-contrast predictors to predict performance. We do so by extending the GLM presented in the previous section to include these predictors. Specifically, here we fit a multilevel logistic regression to all the data (including all contrast band conditions) and include eye movement and image features in the same model. It is unclear how best to incorporate these features into the nonlinear transducer models (a point we return to in the Discussion), but in the GLM framework these are simply included as additional predictors with additional coefficients in the linear predictor (Equation 5). While this is certainly an incorrect oversimplification, we include it here to encourage further development of models to improve predictive performance and explanatory power in data sets like ours.

Method

The predictors we consider here are the pedestal contrast (band energy at the target location), the increment contrast, the geometric invariant K calculated at a midrange spatial scale and across a midrange temporal window (the 1.5–3 cpd condition in Figure 6), the cumulative eye-movement distance during target presentation (Figure 4C), and the spatial band of the target (treated here as a factor, which therefore has five levels). In addition, the model contains an intercept term (which, when the predictors are normalized, is the intercept for the first level of target spatial band). These features yield 10 subject-level regression coefficients, of which four are the slopes of continuous covariates (pedestal, increment, invariant K, and cumulative eye-movement distance). These continuous predictors are first converted to log units as before (replacing any 0 values with the minimum nonzero value), then normalized into z-scores to make the coefficients more interpretable. The remaining coefficients are offsets that cause the regression surface to shift without changing its shape.

In addition to these 10 coefficients for each subject, we also estimated a population mean and deviation parameter for each regression coefficient. Since the subject-level parameters are determined by these population coefficients, the model will incorporate partial pooling (shrinkage) between subject estimates as in the multilevel model above. Details of the fitting of this model are provided in the Appendix.

This model was fitted to a randomly selected subset of 80% of the trials. The remaining 20% of trials were used as a test set to provide a measure of out-of-sample prediction performance. We did not do full cross validation as in the previous section, since this model takes longer to fit than those considered in that section.

Results

Some key results of this model fitting are shown in Figure 13. Figure 13A shows the posterior distributions of the population-level coefficients for the continuous covariates. As expected and as in the simpler GLM before, the probability of success decreases as the pedestal contrast increases (masking) and increases strongly as increment contrast increases (the signal strength of the target increases). As expected from the cumulative eye-movement plot in Figure 4C, more eye movement during the target presentation is correlated with a lower probability of success on a trial. Finally, and unexpectedly, the coefficient for invariant K is negative, indicating that as invariant K increases (as the target location grows in edge density), the probability of success decreases. This relationship has the opposite sign to the positive relationship observed for the same data (the 1.5–3 cpd condition) in Figure 6. This is because here the pedestal contrast is also included in the model, whereas Figure 6 considers only the univariate relationship between invariant K and the response. We return to this point in the Discussion.

Figure 13B shows the population-level regression intercepts for the spatial frequency conditions. The 0.375–0.75 coefficient is the intercept in the model, since this spatial frequency acts as the reference level for the dummy coding of target spatial frequency. In this case, the model intercept corresponds to the value of the linear predictor in the 0.375–0.75 condition when all covariates are at their mean value. The other coefficients of the spatial frequency factor levels are offsets to this intercept. What we plot in Figure 13B are not the raw coefficients but the intercepts, calculated by adding the offsets for the factor levels to the intercept for the model. Higher values of the intercepts correspond to better performance levels, taken when the covariates are at their mean. Interestingly, sensitivity as a function of spatial frequency shows tuning, in that it peaks broadly at 0.75–3 cpd. This fits well with peak sensitivities measured in more standard contrast detection and discrimination paradigms. The variance on the estimates is so high because these are the population-level estimates, and so they are effectively based on only five data points (the subjects in the experiment). To gain more precise estimates of the population-level effects, we would need to test more subjects or impose stronger priors (i.e., assume that members of the population are very similar).

In Figure 13C, we show the predicted proportion correct as a function of the pedestal and increment contrasts as in the previous section, for subject S1 in two spatial frequency conditions. The iso-performance contours have the same slope in both spatial frequency conditions, since the model structure does not allow the slopes to vary. The effect of the differing intercepts is evident in the downward shift of the iso-performance contours in the 0.75–1.5 condition compared to the 0.375–0.75 condition. That is, less increment contrast is required to elicit the same level of performance in the higher spatial frequency condition.

Finally, Figure 13D shows a similar plot but for pedestal contrast versus edge density (geometric invariant K). Performance decreases as pedestal increases, and decreases as edge density increases. This provides another view onto the model surface, which we return to in the Discussion.

As already noted, to examine the predictive performance of this expanded GLM we fitted the model to a random 80% of the trials and then tested its performance on the remaining 20%. For comparison, we did the same for the single-level GLM (i.e., with no population level and therefore no shrinkage) fitted via maximum-likelihood methods. The results are shown in Figure 14. While the multilevel GLM does outperform the single-level comparison, the difference is negligible. We provide these values here mainly to facilitate comparison for future researchers who may wish to compare other models to our simplistic GLM.

Discussion

This article reports human sensitivity to contrast increments presented in natural-image movies in a gaze-contingent paradigm. The contrast increments were localized in both space and spatial frequency. We have presented three analyses that provide a starting point for exploration of this complex data set. We now summarize the key results from each analysis, consider several aspects of the results in greater detail, and provide discussion on related points.

Exploratory data analysis

In Results Part I, we showed that performance increases as a function of the magnitude of the contrast increment and that the strength of this relationship depends on the spatial scale of the increment (Figure 3). Specifically, observers were most sensitive to midrange spatial frequencies between 0.75 and 3 cpd, consistent with peak contrast sensitivity measured in simple stimuli (Campbell & Robson, 1968). Our analysis of an extended logistic regression model supported this impression more quantitatively (Figure 13B).

Sensitivity also depends on the direction, timing, and amplitude of eye movements around the time of target presentation. Saccades initiated after the target onset are biased in the direction of the target (Figure 4A),

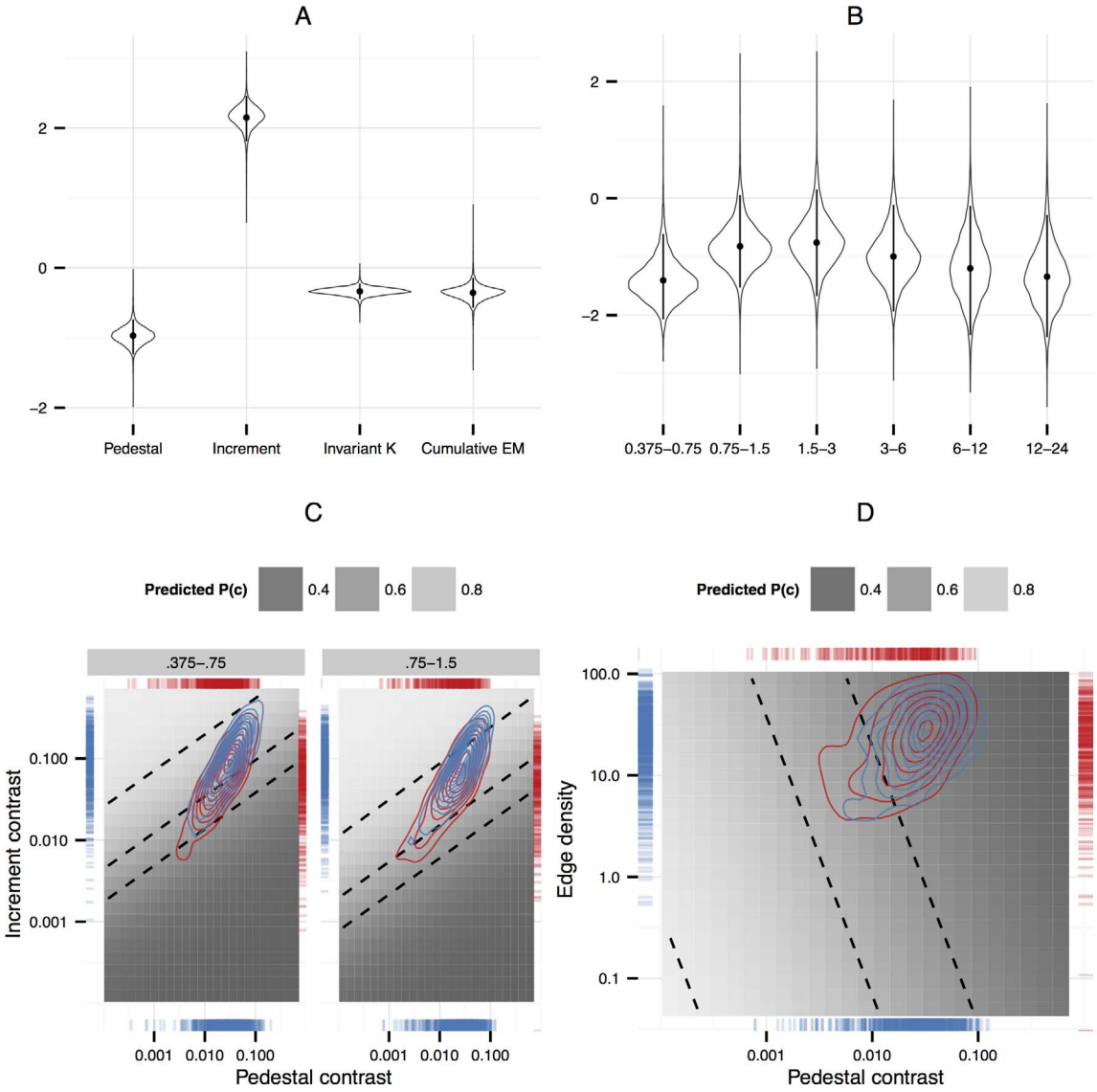

Figure 13. Results for an expanded multilevel GLM fitted to all the data in the experiment. This model includes both eye movement and local image-structure predictors alongside the pedestal and increment contrasts. (A) Posterior distributions of the population-level means of the coefficients for the four continuous predictor variables (covariates). As in Figure 8, kernel density estimates are shown as violin plots; points show the mean of the samples, and error bars give the 95% highest density interval. (B) Posterior distributions of the intercepts corresponding to the different target spatial frequency conditions, calculated by adding the offset coefficient of each factor level to the model intercept. A tuning can be seen in the data, with sensitivity peaking around 1.5–3 cpd in agreement with the typical contrast sensitivity function. (C) Predicted proportion correct as a function of the pedestal and increment contrast for the 0.375–0.75 and 1.5–3 cpd conditions, for subject S1. See Figure 8C for a description of the plot objects. Since spatial frequency is modeled here as an offset with no slope change, the model iso-performance contours shift up and down without changing shape or angle. (D) Predicted proportion correct as a function of the pedestal contrast and the value of geometric invariant *K* (edge density). As edge density increases, the slope of the iso-performance contours decreases.

and performance improves when these saccades are the same size and at latency of the peak of the target contrast (Figure 4D). Performance generally decreases the more the eyes move during the trial (Figure 4C), consistent with our previous observation (Dorr & Bex, 2013). The timing of the subsequent eye movement relative to the target onset appears to have a long-lasting relationship with performance, in that sensitivity declined when saccades were initiated 600 ms after target onset and improved again for trials where saccades were initiated 1200 ms after the target onset. These effects are on much longer timescales than typically reported for peri-saccadic perception, and therefore are likely to represent the influence of higher

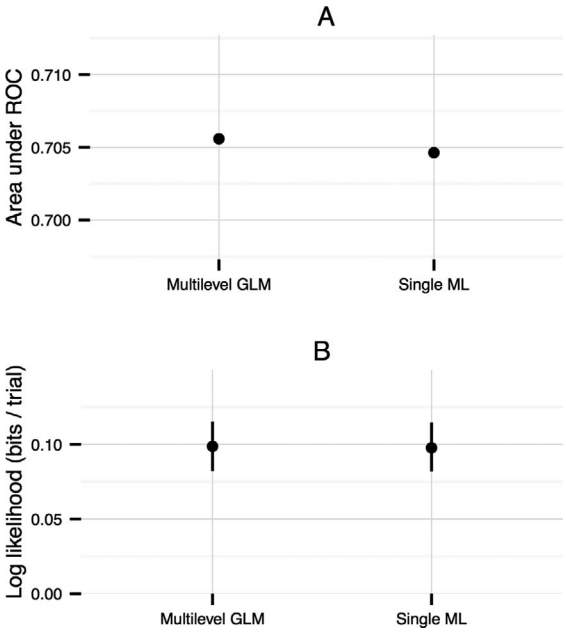

Figure 14. Predictive performance for the expanded multilevel GLM and for the same model fitted separately to each subject using maximum-likelihood methods. (A) The area under the ROC curve (higher is better, chance is 0.5). (B) The average log likelihood (bits per trial, relative to the baseline model). Error bars show 95% confidence intervals derived from 5,000 bootstrap samples. Note that it is not appropriate to compare these numbers to those in Figure 12, since those results are fitted to only a subset of this data set.

level effects such as geotopic mislocalizations of the target (Dorr & Bex, 2013).

The image features at the target location (geometric invariants; loosely, a measure of edge density) were also related to performance (Figure 6), but the positive slope of these relationships may be driven largely by rare low-valued trials, and this relationship changes when considered in conjunction with contrast (see later). We also confirm previous research (Balboa & Grzywacz, 2000; Field, 1987; Mante et al., 2005; Simoncelli, 1997; Zetzsche et al., 1993) showing that the contrast energies in spatial bands of naturalistic images are highly correlated with one another (Figure 2).

Nonlinear transducer models and GLMs

In Results Part II we fitted a subset of the data (the 1.5–3 cpd condition) with a four-parameter nonlinear transducer model and with an atheoretical logistic regression model. When the nonlinear transducer model was fitted with relatively weak prior assumptions about the model parameters (Transducer A), the fitted model produced a posterior distribution with multiple bimodalities (Figure 8) and extreme correlations between the model parameters (Figure 9). Moreover, the fitted posterior means for the model parameters were quite different from the model parameters typically found in studies that employ nonlinear transducer models. We return to this point later. Constraining the model parameters by imposing strong prior distributions based on typical values from the literature led to more reasonable posterior distributions (Transducer B) but produced relatively poor fits to the data (Figure 8D). This was confirmed by comparing the predictive performance of the models using cross validation (Figure 12). In contrast, the GLM parameters were relatively well constrained by the data (Figure 10), the posteriors more sensible (albeit still correlated; Figure 11), and the predictions equivalent to those of the nonlinear Transducer A model (Figure 12).

The four-parameter nonlinear transducer model has been applied with success to many psychophysical data sets (e.g., Bex et al., 2007; Foley, 1994; Haun & Essock, 2010; Haun & Peli, 2013; Holmes & Meese, 2004; Kwon et al., 2008; Meese, Georgeson, & Baker, 2006; Meese & Holmes, 2002; Tolhurst & Tadmor, 1997). It is theoretically appealing because it is a mechanistic model of perceptual processing and can be related to the response profiles of neuronal populations (Goris et al., 2013; Kwon et al., 2008). However, its parameters are quite unconstrained by our data. Our Bayesian analysis shows that the parameters are highly correlated with one another and that the posterior is highly non-Gaussian. Because of this, any interpretation of the contribution of individual model parameters in determining task performance will be heavily dependent on the prior distributions used. As an example from our dataset, if the parameter p is strongly constrained to be around 2, then the *rmax* parameter will not be much greater than 9 (see Figure 9, top right panel). This result could be anticipated from the dependencies between parameters in the model (Haun, 2009; Yu et al., 2003). In additional testing using simulated data and maximum-likelihood fitting (available upon request), we found that wildly different combinations of these four parameters can lead to similar likelihood estimates. Unless parameters are regularized and constrained using prior information, model parameters could be unstable but produce little variation in overall predictive performance: The unconstrained models are not uniquely identifiable from our data.

Is this identifiability problem true for all contrast discrimination datasets? Wichmann (1999) conducted an extensive investigation of several variants of contrast-processing models, including energy detection, nonlinear transduction, and divisive gain control. All parameters in the models could be identified with low variance by fitting to an extensive dataset of classical two-alternative forced-choice contrast discrimination

data (see also Dold, 2012). Furthermore, differentiation of different nested contrast-processing models was achieved based on the Akaike information criterion (Akaike, 1974). It is therefore not always the case that the parameters of nonlinear contrast-processing models cannot be constrained by data. Nevertheless, we believe that our Bayesian analysis of one variant of such a model could encourage future researchers to consider the interdependence of the model parameters and the degree to which they can be constrained by data. We speculate that ours is not the only contrast discrimination data set resulting in a relatively flat likelihood surface for these models.

Interestingly, the slope of psychometric functions for detection are steeper than those for discrimination when the pedestal contrast is at the trough of the dipper (if performance is plotted against Δc; Nachmias & Sansbury, 1974; see also Foley & Legge, 1981; Wichmann, 1999), meaning that the shape of the TvC function depends on the performance level defined as threshold (see also García-Pérez & Alcalá-Quintana, 2007). Wichmann (1999) suggests that differentiating the models depended on using information over the entire psychometric function (see also Green, 1960). This result implies that researchers seeking to compare models of contrast processing should collect full psychometric functions for each pedestal contrast, rather than relying on adaptive methods that seek to estimate only the threshold (and do not constrain the slope).

As discussed in the Introduction, contrast gain control plays a critical role in contrast processing for all but the simplest stimulus conditions (Bex et al., 2007; Bex et al., 2009; Foley, 1994; Geisler & Albrecht, 1992; Haun & Essock, 2010; Haun & Peli, 2013; Heeger, 1992; Holmes & Meese, 2004; Meese & Holmes, 2002, 2007; Morrone et al., 1982). It is therefore unsurprising that the four-parameter contrast-processing model we use in this article does a poor job of fitting the data—it does not explicitly include any cross-scale gain control, nor does it consider orientation or any temporal information. However, given that even this four-parameter model was not constrained by our data, adding more parameters (such as gain-control weights for each scale) without including strong priors will make the models even harder to constrain. This is what we found when we attempted to fit some such models in pilot analyses, and so we chose to present only the simplest version here.

GLM incorporating other predictors

One long-term aim of modeling the early visual system might be to account for results in an experiment like ours. Such a model would necessarily combine a variety of factors that contribute to the observers' behavior, including the experimental manipulations (contrast increment), image content, the observers' eye movements, and even factors such as the high-level scene content (objects, faces). Results Part III presents a rudimentary example of such a combined model. We extended the logistic GLM to include eye movement and image properties, and also allowed the intercept (threshold) to vary with spatial frequency; this extended model was fitted to the entire data set.

Like the GLM in the previous section, this model was relatively well constrained (Figure 13). The thresholds changed with the target spatial band in a manner consistent with the typical contrast sensitivity function, with a peak approximately in the 0.75–3 cpd range (Figure 13B). An additional interesting aspect of this model concerns the influence of image structure on performance. When considered in isolation in Results Part I, geometric invariant K (loosely, a measure of edge density) was positively correlated with task performance: As edge density increased, so did performance (Figure 6). However, when this predictor was included in the expanded GLM, its coefficient became reliably negative. That is, as edge density increased, performance decreased.

Why does this reversal of sign occur? In the expanded GLM, the variability in performance associated with contrast at the target location and the magnitude of the increment can be taken into account by including both the pedestal contrast and the increment contrast in the model. That is, as edge density increases, performance gets worse *at a given level of pedestal contrast*. This result corroborates that shown by Bex et al. (2009), who found that threshold contrast for increment detection in static natural images was higher (i.e., sensitivity decreased) as the local edge density increased. Note however that making this comparison must be treated cautiously, since here the pedestal contrast and the local edge density are correlated. Trials in which the local edge density was low also tended to have lower pedestal contrasts. Therefore, differences in these estimates may be driven in part by a lack of data at high pedestal contrasts when edge density is low, and vice versa. Similarly, it is possible that edge density is related to the correlation in contrast across spatial scales, in that image patches with more edges show higher cross-scale contrast correlations. In a gain-control model, higher cross-scale contrast correlations would produce stronger masking, potentially accounting for the effect of edge density here.[5]

As mentioned before, a hypothetically complete model of early visual processing should account for all the relationships observed in our exploratory analysis. The GLM we present as a first step towards this goal is therefore certainly wrong. It treats all relationships as linear (in the logit). It fits one slope to pedestal contrast and one to increment contrast for the entire data set

(only the intercept varies across spatial frequency conditions). From Figure 3 we know that the slope *does* change across spatial frequencies. A more complete model would account for this. The GLM also does not include many of the predictors we found to influence performance, such as the direction and timing of saccades relative to the target (Figures 4D and 5). More fundamentally, the GLM is an atheoretical model that makes only indirect assumptions about mechanisms of the early visual system and takes preprocessed feature vectors rather than image sequences as its input. On the other hand, GLMs can be extended to a multilevel framework (with a population level over subjects, as we have done here) relatively uncontroversially, whereas it is less clear how to specify population-level hyperpriors over the parameters of the nonlinear transducer. Overall, we believe it is useful to present the GLM model here to encourage further exploration and model comparison for this and similar data sets.

Relationship to existing literature on contrast discrimination

A vast number of studies have measured contrast increment detection performance under a variety of experimental conditions. Here we discuss several studies in greater detail.

Henning and Wichmann (2007) measured contrast discrimination performance for sinusoidal signals and pedestals that were temporally interleaved with noise (i.e., the noise mask was presented on every second frame). In broadband noise, performance was worse overall compared to discrimination in no noise, but the dipper shape of the TvC function remained. When the noise contained a 1.5-octave notch around the signal frequency (i.e., had attenuated power at these frequencies), the dipper shape disappeared. The dipper shape returned again if only the low-pass or high-pass borders of the notch stimulus were included. Henning and Wichmann suggest that these results cannot be explained by a single-channel nonlinear transducer model, an uncertainty model, or a standard gain-control model, but instead that the results are consistent with off-frequency looking (assuming the notched noise prevents off-frequency looking).

Goris et al. (2009) account for these results by incorporating linear–nonlinear units whose responses are normalized (gain control). In broadband noise, the dipper remains because on average across trials, the relative activations of the excitatory and inhibitory components of the divisive normalization model are the same. In notched noise, the excitatory component is constant (driven only by the signal plus pedestal), whereas the activity of the inhibitory gain pool is higher and more variable over trials. An updated model enforcing optimal weighting (Goris et al., 2013) does not predict these data, suggesting that human observers use a suboptimal decision rule under these conditions. That is, humans do not combine the output of the channels in a way that will maximize their performance from the available information. There may be an important component of perceptual learning here: Observers may require deterministic input (such as in simple grating experiments) to tune population decoding for a given task. When stimuli are stochastic (as in noise paradigms, and our experiment is an extreme example), observers cannot learn a close-to-optimal readout.[6]

What would the model of Goris et al. (2009) predict for noise whose spectrum is $1/f$ (i.e., closer on average to the natural scenes used in our experiment)? While Goris et al. do not examine responses of their model to $1/f$ noise, the fact that the dipper returns in low-pass noise (Henning & Wichmann, 2007) suggests that $1/f$ noise might also be expected to show a dipper effect, because the responses of channels sensitive to high frequencies are less masked in $1/f$ noise relative to low-frequency channels and so could still be informative about the presence of the signal. This is indeed what Haun and Essock (2010) report: The dipper effect remains in $1/f$ noise, albeit its depth is reduced and the masking part of the function is shallower relative to narrowband noise.

Recently, Alam et al. (2014) measured the contrast threshold required to detect a Gabor-like target (of 3.9 cpd) imposed on a static natural-image background. They found that observers were most sensitive when the Gabor structure was imposed on relatively blank patches, less sensitive for patches containing simple edges, still less sensitive for patches containing dense structure or textures, and least sensitive for dark patches. They applied a contrast-processing model similar to that of Watson and Solomon (1997), including a band-limited channel decomposition and contrast gain control, to predict thresholds. The parameters were drawn from previous research rather than being fitted to the data. The model appears to perform remarkably well over most images—though we would be interested to see it evaluated against a descriptive GLM-like model in terms of likelihood or information gain. The biggest differences between that article and our efforts here are that Alam et al. evaluate detection performance for a target known to the observer (its orientation and spatial frequency), with foveal viewing (a static spatial three-alternative forced-choice paradigm), whose structure is imposed on the image (rather than focusing on existing image structure). Furthermore, in that article the model fits are evaluated against threshold values alone rather than against all individual trials, as we do here. We believe that article provides an excellent complementary data

set to our own, and would be interested to see the extent to which extended contrast-processing models predict both data sets.

Finally, Bradley, Adams, and Geisler (2014) recently presented an impressively general model to predict the contrast-detection thresholds of targets presented in the central 10° of the visual field, on arbitrary grayscale backgrounds (including natural images). As with Alam et al. (2014), we would be interested to see the predictions of this model for the data we present here. However, the model of Bradley et al. (2014) predicts detection thresholds in a task where both the form of the target (e.g., a vertical Gabor) and its location in the scene are known to the observer. These conditions are quite unlike our experiment, where both the target location and its spatial structure are unknown before the trial. Therefore, while their model is designed to account for performance under a wide range of conditions, it would require additional extension to be able to predict performance in our task.

Model estimation and comparison

In this article we employ a Bayesian approach to model fitting, using MCMC methods to estimate the posterior distribution over model parameters. In doing so we attempt to adhere to the general philosophy advocated by Gelman and Shalizi: "*The model*, for a Bayesian, is the combination of the prior distribution and the likelihood, each of which represents some compromise among scientific knowledge, mathematical convenience and computational tractability. ... We make some assumptions, state them clearly, see what they imply, and check the implications" (2012, p. 20). We treat prior distributions not as subjective measures of personal beliefs but as a way to condition and restrict maximum-likelihood estimates similar to, for example, regularized regression (Hastie, Tibshirani, & Friedman, 2009). This approach allowed us to demonstrate the interdependence of the parameters in the models and test the effect of imposing stronger constraints on the parameters within the same modeling framework (see, e.g., Figure 9). We believe the flexibility and computational power of this approach offer many opportunities for useful application in vision science.

We have compared models using cross validation, in which the model is fitted to one part of the data and its predictions for the unseen "test" data are evaluated. Models can then be compared on their relative prediction performance (see, e.g., Figure 12). Other possibilities for model comparison include methods based on overall model likelihood, such as the Akaike information criterion (Akaike, 1974) and the (so-called) Bayesian information criterion (Raftery, 1995; Schwarz, 1978), or Bayesian methods such as Bayes factors (e.g., Kass & Raftery, 1995; M. D. Lee & Wagenmakers, 2014). Likelihood-based methods penalize the overall model likelihood by some measure of the model's complexity (number of free parameters). These criteria therefore trade off model *fit* with *parsimony* and by extension generalizability. A Bayes factor is the ratio of the marginal posterior probability of two models and includes a natural preference for parsimony, since models are penalized for complexity by spreading posterior mass over more parameters (dimensions in the posterior).

We chose to compare models based on cross validation here because while being computationally intensive, it is perhaps the most easily interpreted route to estimating expected predictive accuracy for unseen data (Gelman, Hwang, & Vehtari, 2014), particularly when comparing models of radically different structure. Cross validation also contains an implicit penalty for model complexity: Complex models with too many parameters will fit noise in the training set, leading to poorer performance in the test set (i.e., poor generalizability to new data). Purely likelihood-based methods such as the Akaike information criterion are not ideal for Bayesian methods of model fitting, since the notion of how many parameters are free is not clear (a strong prior distribution can considerably restrict the posterior range of a parameter even though it is still fitted to data). This is particularly true for multilevel models, like the one used in Results Part III. While Bayes factors are theoretically appealing, their actual estimation can be greatly affected by the specification of the priors over the model parameters: Priors that have almost no influence on the posterior of a parameter within a model can have considerable effect on Bayes factors (Gelman & Shalizi, 2012). It is therefore important to ensure that the prior support of all models is equally (un)restrictive for any Bayesian model-comparison attempt. For models with different structures (as in the nonlinear transducer and GLMs reported here), this is difficult to do (but not impossible, by using a sensitivity analysis; M. D. Lee & Wagenmakers, 2014). If one is interested in predictive performance, we find the conceptually straightforward cross-validation approach more appealing.

Caveats and limitations

Perhaps the most important caveat to bear in mind when considering our results is that our task is very different from that of a classical contrast discrimination experiment. In classical situations, the observer usually has full knowledge of the appearance of the target he or she is trying to detect and minimal uncertainty about the target's possible location in space and time. In contrast,

the appearance of the target in our experiment depended on the image content at the target location, and its presentation was both spatially and temporally uncertain. Our observers were able to engage in an attentionally demanding task (watching a television series) in addition to the experimental task (reporting the location of the image modification). The requirement to report the location of the target in retinal coordinates could lead to errors despite observers' having definitely seen the target, depending on the timing and direction of eye movements (geotopic mislocalization; Dorr & Bex, 2013). Performance therefore depends on many extraneous factors whose influence is intentionally minimized in classical experiments, making it difficult to interpret the overall relationship of our results to classical spatial-vision research.

While our analysis in Results Part II primarily considered the contrast at the target location and the size of the increment applied, it is likely that observers also use other cues to detect the target location. For example, eye movements will cause unnatural temporal signals due to the movement of the spatiotemporal contrast envelope across the (normally relatively static) scene. It is therefore possible that observers use a temporal rather than a spatial signal. This particular hypothesis seems unlikely to describe the primary signal observers use, since their performance was better when their eyes remained relatively static (see Figure 4C). Nevertheless, it is worth keeping in mind that cues other than those we have considered here will undoubtedly affect performance.

Third, since our contrast increments were multiplicative, our paradigm cannot measure detection performance. That is, if the target location happens to contain no contrast (i.e., is a uniform field), then there will also be no contrast increment to detect. While we made this design decision to avoid imposing unnatural structure (e.g., a Gabor-like target) on the target movie, it also means that our ability to measure performance in the low-contrast pedestal range is limited. This is likely an important contributing factor to the inability of our data set to constrain the nonlinear transducer model. Note however that even imposing a model shape by setting strong priors on the parameters did not improve (and in fact, worsened) predictive power. Furthermore, large regions of very low contrast (e.g., during scene cuts) occurred rarely in our professionally produced video material (21 trials with exactly zero contrast); additionally, observers typically gaze at image regions with relatively more image structure (Reinagel & Zador, 1999; Vig, Dorr, Martinetz, & Barth, 2012), and thus target patches also typically contain nonzero contrast due to spatial correlations in natural scenes (Simoncelli, 1997; Zetzsche et al., 1993).

While the movies used in our task contain naturalistic structure, they are highly postproduced. This is therefore by no means a perfect representation of the natural input of our observers' visual system. On the other hand, watching television itself can be considered a common and highly practiced behavior. We believe the results are a useful step in the use of more naturalistic stimuli.

Finally, thought must be given to the observer model for our task. In a classical contrast discrimination paradigm, the observer sees one interval containing the pedestal plus the increment and one interval containing the pedestal alone. The observer therefore explicitly compares two (or more, for the *m*-alternative forced-choice case) pieces of sensory evidence. Bex et al. (2007) provide a direct analogue of this scenario for detection in (static) natural scenes. In our analysis, we define the pedestal as the contrast in the unmodified video sequence at the target location. In contrast to the classical case, in our experiment the observer is never shown the pedestal-only interval, and the pedestals in the nontarget locations are not the same as that in the target location. Observers' judgments in our task could be conceived as a comparison to an internal standard for "naturalness." The observer sees four intervals (possible target locations) and compares each one to what he or she expects the natural contrast in that location to be. One decision strategy is to pick the location that most violates the observer's expectation. Since contrast in natural scenes is correlated over space (Simoncelli, 1997; Tkačik et al., 2011; Zetzsche et al., 1993), the expectation could be based on the scene information surrounding the patch: The surrounding image structure may act as a spatially extended pedestal. If this is the case, we think it is reasonable to employ the analyses for classical contrast discrimination experiments as we have done here, even though the observer models are not the same.

Conclusion

We have collected a contrast increment detection data set with gaze-contingent targets in naturalistic movies and summarized the most striking results. By estimating the posterior distribution over model parameters, we show that a version of the standard nonlinear transducer model for contrast discrimination is not well constrained by our data, whereas an atheoretical logistic GLM fares better. We present one way in which models could be extended to include additional noncontrast features.

We believe that mechanistic models of early visual processing are a useful and principled way to test hypotheses about the representations and transformations applied by the visual system. However, they often contain many free parameters, and we believe that

constraining these parameters requires fitting and prediction over a wide range of data sets. The Bayesian methods we have employed here are powerful (allowing for detailed model analysis; see, e.g., Kruschke, 2011; M. D. Lee & Wagenmakers, 2014) and general (seamlessly extendable to a wide range of model classes, such as the multilevel models we use here), and we hope that other researchers will consider using them. To facilitate reuse of our data and analyses, we have made both available for download (see Appendix). In addition, we hope that others will consider comparing the performance of mechanistic models to descriptive, simple, and atheoretical models (such as GLMs) to quantify the information gain provided by mechanistic complexity in a principled way.

Keywords: contrast sensitivity, contrast discrimination, Bayesian statistics, eye movements, modeling

Acknowledgments

The experiment was designed by, and the article written by, all three authors. MD programmed the experiment and collected the data. TSAW analyzed the data. The authors thank Andrew Haun and Felix Wichmann for many useful suggestions after a close reading of an earlier version of this manuscript, and Simon Barthelmé for suggesting we look further into GLMs. Remaining errors and misinterpretations are our own. TSAW was supported by the National Health and Medical Research Council of Australia (NHMRC training fellowship 634560) and an Alexander von Humboldt Postdoctoral Fellowship. PJB was supported by NIH grants EY019281 and EY018664. MD was supported by the Elite Network Bavaria, funded by the Bavarian State Ministry for Research and Education.

*TSAW and MD contributed equally to this article.
Commercial relationships: none.
Corresponding author: Thomas S. A. Wallis.
Email: thomas.wallis@uni-tuebingen.de.
Address: Department of Computer Science and Werner Reichardt Centre for Integrative Neuroscience, University of Tübingen, Tübingen, Germany.

Footnotes

[1] Note that these are standard deviations, so in the contrast-processing model presented in Equation 4, $p = 2$ is a type of energy detector.

[2] Note that this means we are estimating more parameters (the mean and variance for each predictor), so it is not the case that this model has only three parameters where the nonlinear transducers have four. Nevertheless, we believe the gain in interpretability is worth the extra complexity, which is standard practice when employing GLMs.

[3] In a pilot model fit, we tested an interaction term and found that its coefficient did not differ credibly from 0, so we exclude it here for simplicity.

[4] For the GLMs, the normalization of the predictors (into z-scores) was also cross validated by using the mean and standard deviation of the training set to normalize the test set.

[5] We credit Andrew Haun for this suggestion.

[6] We credit Felix Wichmann for this suggestion.

References

Adelson, E. H., & Burt, P. J. (1981). Image data compression with the Laplacian pyramid. In J. Editor (Ed.), *Proceedings of the Conference on Pattern Recognition and Image Processing* (pp. 218–223). Los Angeles: IEEE Computer Society Press.

Agostinelli, C., & Lund, U. (2013). R package circular: Circular Statistics (version 0.4–7) [Computer software]. Retrieved from https://r-forge.r-project.org/projects/circular/.

Akaike, H. (1974). A new look at the statistical model identification. *IEEE Transactions on Automatic Control, 19*(6), 716–723.

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, doi:10.1167/14.8.22. [PubMed] [Article]

Balboa, R. M., & Grzywacz, N. M. (2000). Occlusions and their relationship with the distribution of contrasts in natural images. *Vision Research, 40*(19), 2661–2669.

Balboa, R. M., & Grzywacz, N. M. (2003). Power spectra and distribution of contrasts of natural images from different habitats. *Vision Research, 43*(24), 2527–2537.

Barth, E., & Watson, A. (2000). A geometric framework for nonlinear visual coding. *Optics Express, 7*(4), 155–165.

Bex, P. J. (2010). (In) sensitivity to spatial distortion in natural scenes. *Journal of Vision, 10*(2):23, 1–15, doi:10.1167/10.2.23. [PubMed] [Article]

Bex, P. J., & Makous, W. (2002). Spatial frequency, phase, and the contrast of natural images. *Journal of the Optical Society of America A, 19*(6), 1096–1106.

Bex, P. J., Mareschal, I., & Dakin, S. C. (2007).

Contrast gain control in natural scenes. *Journal of Vision, 7*(11):12, 1–12, doi: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, doi:10.1167/9.10.1. [PubMed] [Article]

Blakemore, C., & Campbell, F. W. (1969). On the existence of neurones in the human visual system selectively sensitive to the orientation and size of retinal images. *The Journal of Physiology, 203,* 237–260.

Bradley, C., Abrams, J., & Geisler, W. S. (2014). Retina-V1 model of detectability across the visual field. *Journal of Vision, 14*(12):22, 1–22, doi:10.1167/14.12.22. [PubMed] [Article]

Caelli, T., & Moraglia, G. (1986). On the detection of signals embedded in natural scenes. *Perception & Psychophysics, 39*(2), 87–95.

Campbell, F. W., & Robson, J. G. (1968). Application of Fourier analysis to the visibility of gratings. *The Journal of Physiology, 197*(3), 551–566.

Cannon, M. W., & Fullenkamp, S. C. (1991). A transducer model for contrast perception. *Vision Research, 31*(6), 983–998.

Chandler, D. M., Gaubatz, M. D., & Hemami, S. S. (2009). A patch-based structural masking model with an application to compression. *EURASIP Journal on Image and Video Processing, 2009,* 1–22.

David, S. V., Vinje, W. E., & Gallant, J. L. (2004). Natural stimulus statistics alter the receptive field structure of v1 neurons. *Journal of Neuroscience, 24*(31), 6991–7006.

Deubel, H., & Schneider, W. X. (1996). Saccade target selection and object recognition: Evidence for a common attentional mechanism. *Vision Research, 36*(12), 1827–1837.

Dold, H. M. H. (2012). *On modeling data from visual psychophysics: A Bayesian graphical model approach* (Doctoral dissertation). Technische Universitaet Berlin.

Dorr, M., & Bex, P. J. (2013). Peri-saccadic natural vision. *Journal of Neuroscience, 33*(3), 1211–1217.

Dorr, M., Martinetz, T., Gegenfurtner, K., & Barth, E. (2010). Variability of eye movements when viewing dynamic natural scenes. *Journal of Vision, 10*(10): 28, 1–17, doi:10.1167/10.10.28. [PubMed] [Article]

Eckstein, M. P., Ahumada, A. J., & Watson, A. B. (1997). Visual signal detection in structured backgrounds. II. Effects of contrast gain control, background variations, and white noise. *Journal of the Optical Society of America A, 14*(9), 2406–2419.

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.

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.

Foley, J. M., & Legge, G. E. (1981). Contrast detection and near-threshold discrimination in human vision. *Vision Research, 21*(7), 1041–1053.

Frazor, R. A., & Geisler, W. S. (2006). Local luminance and contrast in natural images. *Vision Research, 46*(10), 1585–1598.

Freeman, J., & Simoncelli, E. P. (2011). Metamers of the ventral stream. *Nature Neuroscience, 14*(9), 1195–1201.

García-Pérez, M. A., & Alcalá-Quintana, R. (2007). The transducer model for contrast detection and discrimination: Formal relations, implications, and an empirical test. *Spatial Vision, 20*(1–2), 5–43.

Geisler, W. S. (2008). Visual perception and the statistical properties of natural scenes. *Annual Review of Psychology, 59*(1), 167–192.

Geisler, W. S., & Albrecht, D. G. (1992). Cortical neurons: Isolation of contrast gain control. *Vision Research, 32*(8), 1409–1410.

Geisler, W. S., Najemnik, J., & Ing, A. D. (2009). Optimal stimulus encoders for natural tasks. *Journal of Vision, 9*(13):17, 1–16, doi:10.1167/9.13.17. [PubMed] [Article]

Geisler, W. S., & Perry, J. S. (2009). Contour statistics in natural images: Grouping across occlusions. *Visual Neuroscience, 26*(1), 109–121.

Gelman, A. (2006). Multilevel (hierarchical) modeling: What it can and cannot do. *Technometrics, 48*(3), 432–435.

Gelman, A., & Hill, J. (2007). *Data analysis using regression and multilevel/hierarchical models*. New York: Cambridge University Press.

Gelman, A., Hwang, J., & Vehtari, A. (2014). Understanding predictive information criteria for Bayesian models. *Statistics and Computing, 24*(6), 997–1016.

Gelman, A., & Rubin, D. B. (1992). Inference from iterative simulation using multiple sequences. *Statistical Science, 7*(4), 457–472.

Gelman, A., & Shalizi, C. R. (2012). Philosophy and the practice of Bayesian statistics. *British Journal of*

Mathematical and Statistical Psychology, 66(1), 8–38.

Goris, R. L. T., Putzeys, T., Wagemans, J., & Wichmann, F. A. (2013). A neural population model for visual pattern detection. *Psychological Review, 120*(3), 472–496.

Goris, R., Wichmann, F. A., & Henning, G. (2009). A neurophysiologically plausible population code model for human contrast discrimination. *Journal of Vision, 9*(7):15, 1–22, doi:10.1167/9.7.15. [PubMed] [Article]

Graham, N., & Nachmias, J. (1971). Detection of grating patterns containing two spatial frequencies: A comparison of single-channel and multiple-channels models. *Vision Research, 11*(3), 251–259.

Graham, N., Robson, J. G., & Nachmias, J. (1978). Grating summation in fovea and periphery. *Vision Research, 18*(7), 815–825.

Green, D. M. (1960). Psychoacoustics and detection theory. *The Journal of the Acoustical Society of America, 32*(10), 1189–1203.

Hacker, M. J., & Ratcliff, R. (1979). A revised table of d' for M-alternative forced choice. *Attention, Perception & Psychophysics, 26*(2), 168–170.

Hansen, B. C., & Hess, R. F. (2012). On the effectiveness of noise masks: Naturalistic vs. unnaturalistic image statistics. *Vision Research, 60,* 101–113.

Hastie, T., Tibshirani, R., & Friedman, J. (2009). *The elements of statistical learning*. New York: Springer.

Haun, A. M. (2009). *Contrast sensitivity in 1/f noise* (Doctoral dissertation). University of Louisville, Kentucky.

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, doi: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, doi:10.1167/13.13.3. [PubMed] [Article]

Heeger, D. J. (1992). Normalization of cell responses in cat striate cortex. *Visual Neuroscience, 9*(2), 181–197.

Henning, G. B., & Wichmann, F. A. (2007). Some observations on the pedestal effect. *Journal of Vision, 7*(1):3, 1–15, doi:10.1167/7.1.3. [PubMed] [Article]

Hering, E. (1879). Über die Muskelgeräusche des Auges. *Sitzberichte der kaiserlichen Akademie der Wissenschaften in Wien. Mathematisch-naturwissenschaftliche Klasse*, *79*, 137–154.

Hoffman, M. D., & Gelman, A. (2014). The No-U-Turn Sampler: Adaptively setting path lengths in Hamiltonian Monte Carlo. *Journal of Machine Learning Research, 15*(Apr), 1593–1623.

Holmes, D. J., & Meese, T. S. (2004). Grating and plaid masks indicate linear summation in a contrast gain pool. *Journal of Vision, 4*(12):7, 1080–1089, doi:10.1167/4.12.7. [PubMed] [Article]

Kass, R. E., & Raftery, A. E. (1995). Bayes factors. *Journal of the American Statistical Association, 90*(430), 773–795.

Kelly, D. H. (1984). Retinal inhomogeneity. I. Spatiotemporal contrast sensitivity. *Journal of the Optical Society of America A, 1*(1), 107–113.

Knoblauch, K. (2014). psyphy: Functions for analyzing psychophysical data in R [Computer software]. Location: Publisher.

Knoblauch, K., & Maloney, L. T. (2012). *Modeling psychophysical data in R*. New York: Springer.

Kruschke, J. K. (2011). *Doing Bayesian data analysis*. Burlington, MA: Academic Press/Elsevier.

Kwon, M., Legge, G. E., Fang, F., Cheong, A., & He, S. (2008). Adaptive changes in visual cortex following prolonged contrast reduction. *Journal of Vision, 9*(2):20, 1–16, doi:10.1167/9.2.20. [PubMed] [Article]

Lee, A. B., Mumford, D., & Huang, J. (2001). Occlusion models for natural images: A statistical study of a scale-invariant dead leaves model. *International Journal of Computer Vision, 41*(1–2), 35–59.

Lee, M. D., & Wagenmakers, E.-J. (2014). *Bayesian cognitive modeling: A practical course*. Cambridge, UK: Cambridge University Press.

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

Lennie, P., & Movshon, J. A. (2005). Coding of color and form in the geniculostriate visual pathway (invited review). *Journal of the Optical Society of America A, 22*(10), 2013–2033.

Mante, V., Frazor, R. A., Bonin, V., Geisler, W. S., & Carandini, M. (2005). Independence of luminance and contrast in natural scenes and in the early visual system. *Nature Neuroscience, 8*(12), 1690–1697.

Mark, W. R., Steven, R., Kurt, G., Mark, A., & Kilgard, J. (2003). Cg: A system for programming graphics hardware in a c-like language. *ACM Transactions on Graphics*, *22*, 896–907.

Meese, T. S., & Georgeson, M. A. (2005). Carving up the patchwise transform: Towards a filter combi-

nation model for spatial vision. In S. P. Shohov (Ed.), *Advances in Psychology Research, Vol. 34* (pp. 51–88). New York: Nova Science Publishers.

Meese, T., Georgeson, M. A., & Baker, D. H. (2006). Binocular contrast vision at and above threshold. *Journal of Vision, 6*(11):7, 1224–1243, doi:10.1167/6.11.7. [PubMed] [Article]

Meese, T. S., & Holmes, D. J. (2002). Adaptation and gain pool summation: Alternative models and masking data. *Vision Research, 42*(9), 1113–1125.

Meese, T. S., & Holmes, D. J. (2007). Spatial and temporal dependencies of cross-orientation suppression in human vision. *Proceedings of the Royal Society B: Biological Sciences, 274*(1606), 127–136.

Morrone, M. C., Burr, D. C., & Maffei, L. (1982). Functional implications of cross-orientation inhibition of cortical visual cells. I. Neurophysiological evidence. *Proceedings of the Royal Society of London: B, 216*(1204), 335–354.

Moscatelli, A., Mezzetti, M., & Lacquaniti, F. (2012). Modeling psychophysical data at the population-level: The generalized linear mixed model. *Journal of Vision, 12*(11):26, 1–17, doi:10.1167/12.11.26. [PubMed] [Article]

Nachmias, J., & Sansbury, R. (1974). Letter: Grating contrast: Discrimination may be better than detection. *Vision Research, 14*(10), 1039–1042.

Oppenheim, A. V., & Lim, J. S. (1981). The importance of phase in signals. *Proceedings of the IEEE, 69,* 529–541.

Papaspiliopoulos, O., Roberts, G. O., & Sköld, M. (2007). A general framework for the parametrization of hierarchical models. *Statistical Science, 22*(1), 59–73, http://doi.org/10.1214/088342307000000014.

Peli, E. (1990). Contrast in complex images. *Journal of the Optical Society of America A, 7*(10), 2032–2040.

Poynton, C. (2003). *Digital video and HDTV*. San Francisco: Morgan Kaufmann.

R Core Development Team. (2013). R: A language and environment for statistical computing [Computer software]. Vienna, Austria: R Foundation for Statistical Computing.

Raftery, A. E. (1995). Bayesian model selection in social research. In P. V. Marsden (Ed.), *Sociological methodology* (pp. 111–196). Cambridge, MA: Blackwell.

Reinagel, P., & Zador, A. M. (1999). Natural scene statistics at the centre of gaze. *Network: Computation in Neural Systems, 10,* 341–350.

Ringach, D. L., Hawken, M. J., & Shapley, R. (1997). Dynamics of orientation tuning in macaque primary visual cortex. *Nature, 387*(6630), 281–284.

Ringach, D. L., Hawken, M. J., & Shapley, R. (2002). Receptive field structure of neurons in monkey primary visual cortex revealed by stimulation with natural image sequences. *Journal of Vision, 2*(1):2, 12–24, doi:10.1167/2.1.2. [PubMed] [Article]

Rovamo, J., Mäkelä, P., Näsänen, R., & Whitaker, D. (1997). Detection of geometric image distortions at various eccentricities. *Investigative Ophthalmology & Visual Science, 38*(5), 1029–1039. [PubMed] [Article]

Ruiz, O., & Paradiso, M. A. (2012). Macaque V1 representations in natural and reduced visual contexts: Spatial and temporal properties and influence of saccadic eye movements. *Journal of Neurophysiology, 108*(1), 324–333.

Sadr, J., & Sinha, P. (2004). Object recognition and Random Image Structure Evolution. *Cognitive Science, 28,* 259–87.

Schwarz, G. (1978). Estimating the dimension of a model. *The Annals of Statistics, 6*(2), 461–464.

Simoncelli, E. P. (1997). Statistical models for images: Compression, restoration and synthesis. In J. Editor (Ed.), *Proceedings of the 31st Asilomar Conference on Signals, Systems and Computers, Vol. 1* (pp. 673–678). Location: IEEE Computer Press.

Sing, T., Sander, O., Beerenwinkel, N., & Lengauer, T. (2005). ROCR: Visualizing classifier performance in R. *Bioinformatics, 21*(20), 3940–3941.

Sinz, F., & Bethge, M. (2013). Temporal adaptation enhances efficient contrast gain control on natural images. *PLoS Computational Biology, 9*(1), e1002889.

Solomon, J. A. (2009). The history of dipper functions. *Attention, Perception & Psychophysics, 71*(3), 435–443.

Stan Development Team. (2014). Stan modeling language users guide and reference manual, version 2.2. Retrieved from http://mc-stan.org/.

Stromeyer, C. F., & Klein, S. (1974). Spatial frequency channels in human vision as asymmetric (edge) mechanisms. *Vision Research, 14*(12), 1409–1420.

Thomson, M. G. (1999). Visual coding and the phase structure of natural scenes. *Network: Computation in Neural Systems, 10*(2), 123–132.

Tkačik, G., Garrigan, P., Ratliff, C., Milcinski, G., Klein, J. M., Seyfarth, L. H., ... Balasubramanian, V. (2011). Natural images from the birthplace of the human eye. *PLoS ONE, 6*(6), e20409.

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.

van Hateren, J. H., & Ruderman, D. L. (1998). Independent component analysis of natural image sequences yields spatio-temporal filters similar to simple cells in primary visual cortex. *Proceedings of the Royal Society of London. Series B: Biological Sciences, 265*(1412), 2315–2320.

Vig, E., Dorr, M., Martinetz, T., & Barth, E. (2012). Intrinsic dimensionality predicts the saliency of natural dynamic scenes. *IEEE Transactions on Pattern Analysis and Machine Intelligence, 34*(6), 1080–1091.

Vogels, R. (1999). Effect of image scrambling on inferior temporal cortical responses. *NeuroReport, 10*(9), 1811–1816.

Wallis, T. S. A., & Bex, P. J. (2012). Image correlates of crowding in natural scenes. *Journal of Vision, 12*(7): 6, 1–19, doi:10.1167/12.7.6. [PubMed] [Article]

Wang, H. X., Freeman, J., Merriam, E. P., Hasson, U., & Heeger, D. J. (2012). Temporal eye movement strategies during naturalistic viewing. *Journal of Vision, 12*(1):16, 1–27, doi:10.1167/12.1.16. [PubMed] [Article]

Watson, A. B., & Ahumada, A. J. (2005). A standard model for foveal detection of spatial contrast. *Journal of Vision, 5*(9):6, 717–740, doi:10.1167/5.9.6. [PubMed] [Article]

Watson, A. B., & Solomon, J. A. (1997). Model of visual contrast gain control and pattern masking. *Journal of the Optical Society of America A, 14*(9), 2379–2391.

Wichmann, F. (1999). *Some aspects of modelling human spatial vision: Contrast discrimination* (Doctoral dissertation). University of Oxford.

Wichmann, F. A., Braun, D. I., & Gegenfurtner, K. R. (2006). Phase noise and the classification of natural images. *Vision Research, 46*(8), 1520–1529.

Wood, S. N. (2011). Fast stable restricted maximum likelihood and marginal likelihood estimation of semiparametric generalized linear models. *Journal of the Royal Statistical Society: Series B (Statistical Methodology), 73*(1), 3–36.

Xavier Fernández i Marín. (2014). ggmcmc: Graphical tools for analyzing Markov Chain Monte Carlo simulations from Bayesian inference. Retrieved from http://xavier-fim.net/packages/ggmcmc.

Yu, C., Klein, S. A., & Levi, D. M. (2003). Cross- and iso-oriented surrounds modulate the contrast response function: The effect of surround contrast. *Journal of Vision, 3*(8):1, 527–530, doi:10.1167/3.8.1. [PubMed] [Article]

Zetzsche, C., & Barth, E. (1990). Fundamental limits of linear filters in the visual processing of two-dimensional signals. *Vision Research, 30*(7), 1111–1117.

Zetzsche, C., Barth, E., & Wegmann, B. (1993). The importance of intrinsically two-dimensional image features in biological vision and picture coding. In A. B. Watson (Ed.), *Digital images and human vision* (pp. 109–138). Cambridge, MA: MIT Press.

Appendix

Here we outline details of our fitting procedures and the structure of the models presented in the article. The analyses were implemented using the free software package R with other free libraries. The code and data set for reproducing all analyses in this article are available from the first author's GitHub page (http://github.com/tomwallis/gcd_contrast). The data set we provide online includes numerous features (e.g., local average pixel intensity, a number of additional geometric invariants, additional eye-movement characteristics) not considered in this article. We encourage interested readers to explore the data set themselves.

Nonlinear transducer model structure

In Results Part II we fitted two versions of a nonlinear transducer model with four parameters for each subject. The two versions differed in their priors. The parameters of the models were p, q, z, and $rmax$ (see Equation 4).

The following priors were used in Transducer A for each subject: We placed priors over the marginal distributions of the parameters (thus assuming that the parameters were independent a priori). All four parameters were given a lower bound of 0. This ensured that the relationship between response and contrast was positive. Parameter z was given a uniform prior bounded [0, 1], and $rmax$ was given a uniform prior bounded [0, 100]. Parameter p was given a Gaussian (normal) prior with mean 2 and standard deviation 1, and parameter q was given a normal prior with mean 0.4 and standard deviation 0.2. That is, the prior variances were set to the mean divided by 2. The means of these parameters were picked to conform to the standards used in the literature to produce the dipper effect, whereas the variance was picked to ensure that the priors were reasonably weakly centered over the means, giving the data the opportunity to influence the model. The lower asymptote γ was fixed at chance performance, 0.25.

In contrast, Transducer B had strong priors over all marginal parameters except z, which retained the same [0, 1] uniform distribution. The parameters were bounded as in Transducer A. The prior over *rmax* was a normal distribution with mean 10 and standard deviation 0.1, the prior over p was normal with mean 2 and standard deviation 0.02, and the prior over q was normal with mean 0.4 and standard deviation 0.004. That is, the prior variances were set to the mean divided by 100, which we intended to strongly constrain the posterior distribution around the mean values. While the value of 100 was picked arbitrarily, the results show that we achieved our intended target.

GLM structure

Two variants of a GLM were fitted in Results Part II. The first was a single-level GLM in which the coefficients for each subject were estimated independently. In this model, each coefficient $\beta_{i,j}$ for subject i and predictor j was bounded to [−5, 5] and given a normal distribution prior with a mean of 0 and a standard deviation of 2. These values were chosen to represent quite weak priors centered over 0 (i.e., no effect). Note that the predictors were standardized (z-transformed) after taking their log, so these coefficients represent the weighting of standardized log units.

The second model included a multilevel structure that related the subject coefficients via a population distribution over the coefficient values. Each subject-level coefficient $\beta_{i,j}$ was unbounded. The population-level mean for each predictor variable j is denoted μ_j, and the population standard deviation for each predictor variable is denoted σ_j. These hyperparameters were bounded between [−5, 5] and [0, 10], respectively. The population means μ_j were given normal distribution priors with a mean of 0 and a standard deviation of 1. The standard deviation σ_j priors were half-Cauchy distributions, with mean 0 and standard deviation 1. These values were picked to make the subject-level estimates close to one another, since the bulk of the prior density is centered over 0 (i.e., no variance between subjects). This represents a conservative assumption, common in psychophysics, that the subjects do not differ greatly. We feel this is appropriate here, since we have only five subjects and the number of trials from each subject varies.

To aid model convergence, we parameterized the subject-level coefficients as a unit-normal offset variable that we then shifted by the population mean and scaled by the population standard deviation to produce the final coefficient estimate (Papaspiliopoulos et al., 2007; Stan Development Team, 2014). That is, each individual subject parameter $\beta_{i,j}$ is given by

$$\beta_{i,j} = \mu_j + \varepsilon_{i,j}\sigma_j \quad \text{(A1)}$$

where $\varepsilon_{i,j}$ is the offset for each subject and each coefficient. Each $\varepsilon_{i,j}$ was given a unit normal prior distribution.

The expanded GLM (Results Part III) had an identical structure to the multilevel GLM from Results Part II, with the exception that we lowered the upper bound on the population standard deviation from 10 to 5 due to initialization problems with the expanded model. This impacts the estimates only slightly, since the posterior mass is concentrated between 0 and 1 for this parameter due to the half-Cauchy prior.

MCMC sampling parameters

Both transducer models were fitted using four independent chains sampled 100,000 times each, of which the first 50,000 samples were treated as a warm-up phase to adapt the sampler. To reduce autocorrelation in the samples and to reduce file size, we saved every 40th sample, producing 1,250 post-warm-up samples per chain to total the 5,000 samples analyzed. These sampling parameters produced good chain mixing and convergence results for both models.

The single-level GLM in Results Part II was fitted using 20,000 samples per chain, of which the first 10,000 were discarded (adaptation) and every eighth sample saved to produced 1,250 final samples per chain. The multilevel GLM was sampled for longer, since this model took longer to converge; the sampling parameters for the multilevel model were the same as for the transducer models. The expanded GLM (Results Part III) was also sampled using these parameters.

The fivefold cross validations reported in Results Part II are based on 50,000 samples (25,000 warm-up), saving every 100th sample to produce 250 samples per chain for a total of 1,000 samples that were used to calculate the posterior mean for prediction.

Approximation to d' function

Evaluating the integral determining proportion correct in a signal detection theory (SDT) framework with four alternatives (Equation 2) is costly to implement in Stan. Instead we approximated this function using a Weibull curve:

$$\text{p(correct)} = \frac{1}{m} + (1 - \frac{1}{m})(1 - \exp(-(\Delta_R/\lambda)^k) \quad \text{(A2)}$$

for scale λ and shape k. These parameters were set to 1.545903 and 1.481270, respectively, by minimizing the

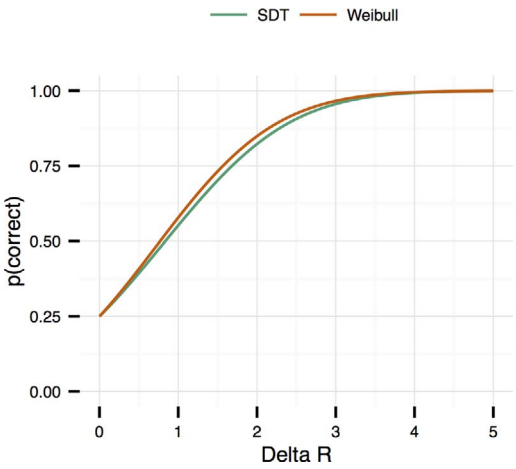

Figure A1. Comparison of the signal detection theory (SDT) link function and our Weibull approximation.

squared difference between the Weibull curve and Equation 2. It offers a reasonable approximation over the range of informative d' values (see Figure A1).

Relationship between pixel intensity and performance

In this article, we have primarily considered the relationship between the contrast in the target location and performance at detecting a contrast increment. However, it is also useful to consider whether the raw pixel intensities (related to luminance) are also associated with performance.

Figure A2 shows the relationship between pixel intensity, contrast, and performance. Figure A2A suggests that pixel intensity and contrast are mostly related by the fact that when luminance is near zero, so too is contrast.

If we consider the relationship between pixel intensity and performance across all trials (solid fits, Figure A2B), it is clear that performance is worse if the average pixel intensity at the target location is relatively low or high. Of course, if the average pixel intensity is near dark or near bright, the contrast will also tend to be low. Can the relationship between pixel intensity and performance be explained by covariance with contrast?

The dashed fit in Figure A2B shows the relationship between luminance and performance for all trials with pedestal contrasts between 0.01 and 0.05. When contrast is held within a nonzero range, the relationship between luminance and performance is largely abolished. Figure A2C shows the relationship between performance and contrast for this subset of the data, demonstrating that contrast and performance are still related.

The luminance at the target location thus does not seem to be uniquely important for predicting performance, beyond its relationship to contrast. This result could perhaps be anticipated from the results of Mante et al., 2005, who show both that luminance and contrast are largely independent in natural scenes, and that luminance gain control is largely independent of contrast gain control.

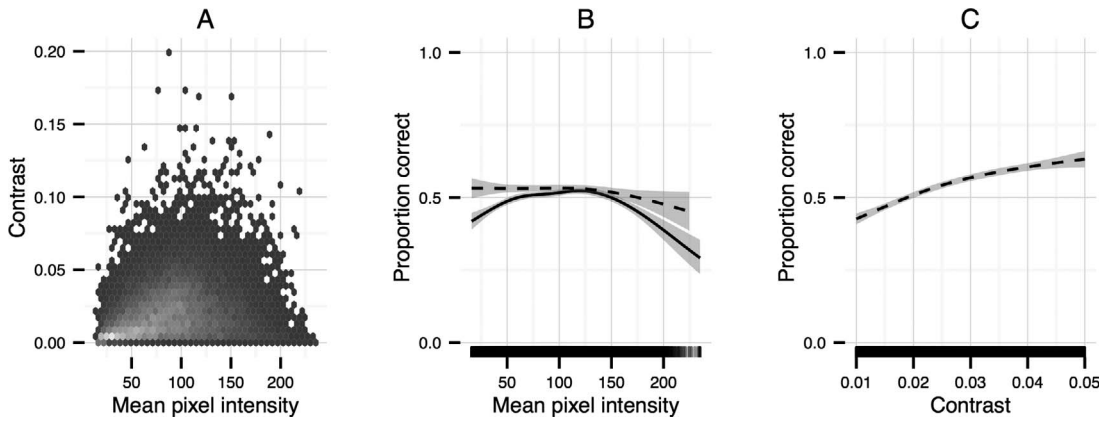

Figure A2. Relationship between pixel intensity and performance. (A) Bivariate relationship between mean pixel intensity at the target location and contrast. Data points are aggregated into hexagonal bins to reduce overplotting; lighter bins represent more samples. The rank-order correlation coefficient is 0.21. (B) Two spline fits between pixel intensity and performance (cubic splines with five knots). The solid line shows the relationship across all trials, whereas the dashed line shows the same fitted to trials with pedestal contrasts between 0.01 and 0.05. When contrast is held within a nonzero range, the relationship between luminance and performance is largely abolished. (C) For the same subset of the data shown by the dashed line in (B), the relationship between contrast and performance remains.

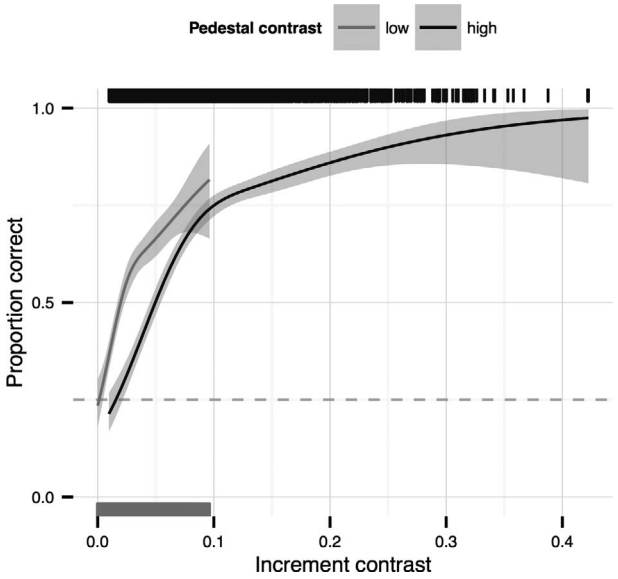

Figure A3. Performance as a function of the increment contrast for observer S1 in the 1.5–3 cpd condition. Target band energy is binned as in Figure 3. Curves and shaded regions show the fits and confidence regions of smoothing splines (cubic spline with five knots, GAM with a logistic link function). Dashes in lower x-axis show the distribution of increment values for the low condition, dashes in upper x-axis show increment values for high pedestal contrasts.

Contrast masking and average performance

Figure 3 shows that performance in trials with high contrast at the target location was better than in trials with low contrast at the target location. Yet our analysis in Results Part II shows stereotypical contrast-masking effects at higher contrast levels, in that contrast increment thresholds are larger as pedestal contrast increases. How do these effects fit together?

The reason for this apparent discrepancy is that Figure 3 shows performance as a function of the multiplication factor. We chose this representation to demonstrate the range of performances elicited by our raw manipulation of the video signal. However, contrast discrimination performance is usually considered in terms of the increment: the *difference* in contrast between the absolute contrast of the target and the pedestal.

To demonstrate the relationship between these representations, we replotted the performance of observer S1 in the 1.5–3 cpd condition as a function of the increment contrast (i.e., pedestal × multiplication factor − pedestal). Figure A3 shows that the high pedestal contrasts have more data points at higher increment values—and consequently higher average performance than the low pedestal contrasts. Where the increments overlap between the pedestal contrast bins, however, performance in the low condition is better than in the high condition. That is, when the pedestal contrast is high, a larger contrast increment is required to reach the same level of performance as when the pedestal is low. This is contrast masking.