Detecting distortions of peripherally-presented letter stimuli under crowded conditions

Thomas S. A. Wallis[1, 2, 3, 7], Saskia Tobias[1], Matthias Bethge[1, 3, 5, 6], and Felix A. Wichmann[1, 3, 4]

[1]Neural Information Processing Group, Faculty of Science, Eberhard Karls Universität Tübingen

[2]Werner Reichardt Center for Integrative Neuroscience, Eberhard Karls Universität Tübingen

[3]Bernstein Center for Computational Neuroscience, Tübingen

[4]Max Planck Institute for Intelligent Systems, Empirical Inference Department, Tübingen

[5]Institute for Theoretical Physics, Eberhard Karls Universität Tübingen

[6]Max Planck Institute for Biological Cybernetics, Tübingen

[7]To whom correspondence should be addressed: thomas.wallis@uni-tuebingen.de

April 12, 2016

Abstract

When visual features in the periphery are close together they become difficult to recognise: *something* is present but it is unclear what. This is called "crowding". Here we investigated sensitivity to features in highly familiar shapes (letters) by applying spatial distortions. In Experiment 1, observers detected which of four peripherally-presented (8 deg of retinal eccentricity) target letters was distorted (spatial 4AFC). The letters were presented either isolated or surrounded by four undistorted flanking letters, and distorted with one of two types of distortion at a range of distortion frequencies and amplitudes. The bandpass noise distortion ("BPN") technique causes spatial distortions in cartesian space, whereas radial frequency distortion ("RF") causes shifts in polar coordinates. Detecting distortions in target letters was more difficult in the presence of flanking letters, consistent with the effect of crowding. The BPN distortion type showed evidence of tuning, with sensitivity to distortions peaking at approximately 6.5 c/deg for unflanked letters. The presence of flanking letters causes this peak to rise to approximately 8.5 c/deg. In contrast to the tuning observed for BPN distortions, RF distortion sensitivity increased as the radial frequency of distortion increased. In a series of follow-up experiments we found that sensitivity to distortions is reduced when flanking letters were also distorted, that this held when observers were required to report which target letter was *un*distorted, and that this held when flanker distortions were always detectable. The perception of geometric distortions in letter stimuli is impaired by visual crowding.

1   Introduction

When a target object (such as a letter) is presented to the peripheral retina flanked by similar non-target objects (other letters), a human observer's ability to discriminate or identify the target object is impaired relative to conditions where no flankers are present. This "crowding" phenomenon (Andriessen and Bouma, 1975; Levi et al., 1985; Greenwood et al., 2009; Bouma, 1970; Parkes et al., 2001; Toet and Levi, 1992; Strasburger, 2014; Herzog et al., 2015; Harrison and Bex, 2015) is characterised by a reduction in sensitivity to peripheral image structure. One way to physically change image structure is to apply spatial distortion, in which the position of local elements (pixels) are perturbed in some fashion (for example, by stretching or shifting). Characterising human sensitivity to spatial distortions is one way to investigate the perceptual encoding of local image structure. For example, showing that perception is invariant to a certain type of distortion (i.e. things look the same whether physically distorted or not) implies that the human visual system does not encode the distortion in question, either directly or indirectly. Arguably, measuring sensitivity to the distortion of highly familiar shapes such as letters (as we do in this paper) allows one to characterise human perception in a more complex task than (for example) grating orientation discrimination, but one that is more tractable from a modelling perspective than (for example) letter identification, which may require a full model of letter encoding. In addition, psychophysical investigation of spatial distortions is relevant to metamorphopsia—the perception of persistent spatial distortions in everyday life— which is commonly associated with retinal diseases that affect the macular (Wiecek et al., 2014).

Human sensitivity to spatial distortions has been investigated previously in images of faces (Spence et al., 2014; Rovamo et al., 1997; Dickinson et al., 2010; Hole et al., 2002) and natural scenes (Kingdom et al., 2007; Bex, 2010). To our knowledge, only one study has assessed the impact of spatial distortion for letter stimuli. Wiecek et al. (2014) had observers identify letters (26-alternative identification task) distorted with bandpass noise distortion (see below) while varying the spatial scale of distortion, the letter size and the viewing distance. Interestingly, they report an interaction between the spatial scale of distortion (CPL; cycles per letter) and viewing distance (changing letter size), such that for small letters (subtending 0.33 degrees of visual angle) performance was worst for coarse-scaled distortions (2.4 CPL), whereas for large letters (5.4 deg) the most detrimental distortion shifted to a finer scale (4 CPL). This result has important implications for patients with metamorphopsia: a stable retinal distortion may affect letter recognition for some letter sizes but not others, influencing acuity assessments using letter charts (a primary outcome measure for clinical vision assessment; Wiecek et al., 2014).

Here we investigate sensitvity to spatial distortions in letters, under crowded (flanked) and uncrowded (unflanked) conditions. Note that our goal here is distinct from that of Wiecek et al. (2014), who measured the impact of distortions on letter identification. We do not measure letter identification here, but instead use letters as a class of relatively simple, artifical, but highly familiar stimuli to investigate sensitivity to the presence of distortion *per se*. We quantify the detectability of two different types of spatial distortion commonly used in the literature (see also Stojanoski and Cusack, 2014, for another distortion not employed here). In bandpass noise distortions (hereafter referred to as *BPN* distortion; Bex, 2010), pixels are warped according to bandpass filtered noise; this ensures that the distortion occurs on a defined and limited spatial scale. In radial frequency distortions (hereafter referred to as *RF* distortion; Wilkinson et al., 1998; Dickinson et al., 2010), the image is warped by modulating the radius (defined from the image centre) according to a sinusoidal function of some frequency. For our purposes they serve to produce two different graded changes in

letter images. A successful model of form discrimination in humans would explain sensitivity to both types of distortion and any dependence on surrounding letters.

2 Experiment 1

2.1 Methods

Stimuli, data and code associated with this paper are available to download from http://dx.doi.org/10.5281/zenodo.48574.

2.1.1 Observers

Five observers with normal or corrected-to-normal vision participated in this experiment: two of the authors, one lab member and two paid observers (10 Euro per hour) who were unaware of the purpose of the study. All of the observers had prior experience with psychophysical experiments and were between 20 an 31 years of age. All experiments conformed to the Declaration of Helsinki.

2.1.2 Apparatus

Stimuli were displayed on a VIEWPixx LCD (VPIXX Technologies; spatial resolution 1920×1200 pixels, temporal resolution 120 Hz). Outside the stimulus image the monitor was set to mean grey. Observers viewed the display from 60 cm (maintained via a chinrest) in a darkened chamber. At this distance, pixels subtended approximately 0.024 degrees on average (41.5 pixels per degree of visual angle). The monitor was carefully linearised (maximum luminance 212 cd/m^2) using a Gamma Scientific S470 Optometer. Stimulus presentation and data collection was controlled via a desktop computer (12 core i7 CPU, AMD HD7970 graphics card) running Kubuntu Linux (14.04 LTS), using the Psychtoolbox Library (Kleiner et al., 2007; Pelli, 1997; Brainard, 1997, version 3.0.11) and our internal iShow library (http://dx.doi.org/10.5281/zenodo.34217) under MATLAB (The Mathworks, Inc., R2013B). Responses were collected using a RESPONSEPixx button box.

2.1.3 Stimuli

The letters stimuli were a subset of the Sloan alphabet (Sloan, 1959), used commonly on acuity charts to measure visual acuity in the clinic. Target letters were always the letters D, H, K and N; flanker letters were always C, O, R, and Z. Letter images were 64×64 pixels. To prevent border artifacts in distortion, each image was padded with white pixels of length 14 at each side, creating 92×92 pixel images. These padded letter images were distorted according to distortion maps generated from the BPN or RF algorithms (see below) in a Python (v2.7.6) environment, using Scipy's `griddata` function with linear 2D interpolation to remap pixels from the original to the distorted image. That is, the distortion map specifies where to move the pixels from the original image; pixel values in intermediate spaces are linearly interpolated from surrounding pixels to produce smooth distortions.

Bandpass Noise (BPN) distortion: Bex (2010, see also (Rovamo et al., 1997; Wiecek et al., 2014)) describes a method for generating spatial distortions that are localised to a particular spatial passband (see Figure 1A–D). Two random 92×92 samples of zero-mean white noise were filtered by a log exponential filter (see Equation 1 in Bex, 2010):
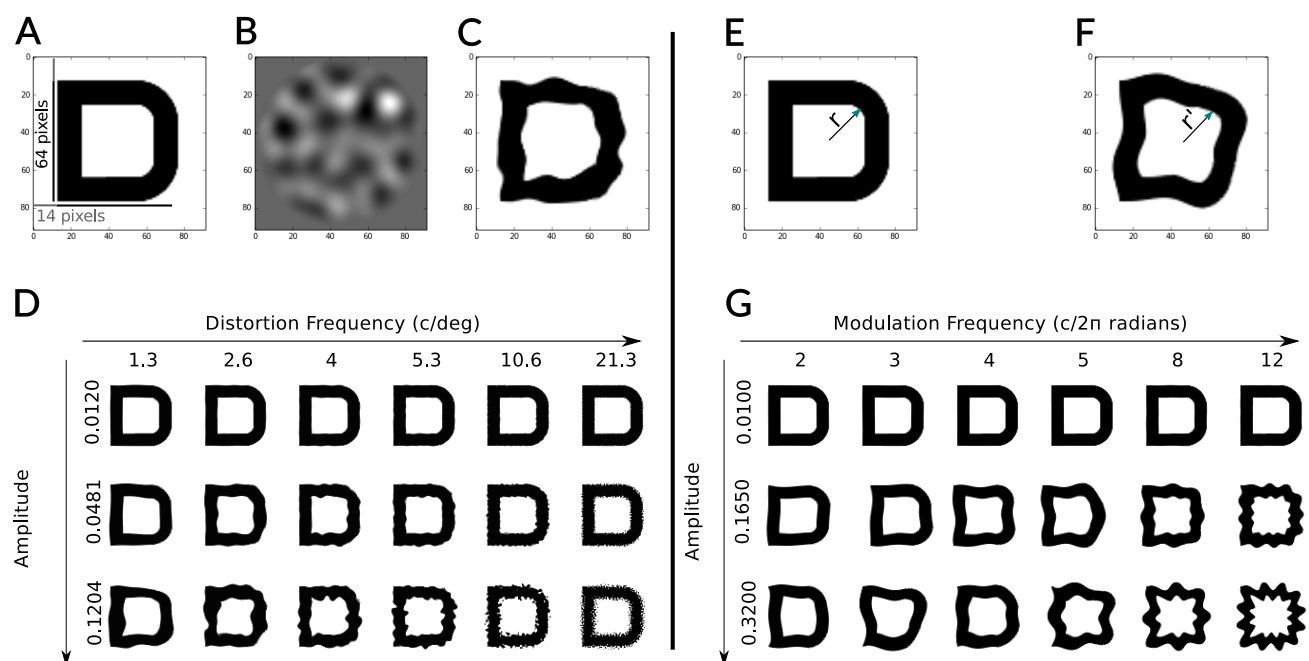
Figure 1: Distortion methods for Bandpass Noise (BPN; A–D) and Radial Frequency (RF; E–G). **A:** A Sloan letter (D) with 14 pixels of white padding. **B:** A sample of bandpass filtered noise, windowed in a circular cosine. Two such noise samples determine the BPN distortion map. **C:** The letter distorted by the BPN technique. **D:** The effects of varying the frequency (columns) and amplitude (rows) of the BPN distortion. **E:** An original letter image, showing the original radius r from the centre to an arbitrary pixel. **F:** RF distortion modulates the radius of every pixel according to a sinusoid, producing a new radius r'. **G:** The effects of varying the frequency (columns) and amplitude (rows) of the RF distortion.

$$A(\omega) \propto \exp\left(-\frac{|\ln(\omega/\omega_{peak})|^3 \ln 2}{(b_{0.5}\ln 2)^3}\right)$$

where ω_{peak} specifies the peak frequency, ω is the spatial frequency and $b_{0.5}$ is the half bandwidth of the filter in octaves. Noise was filtered at one of six peak frequencies (2, 4, 6, 8, 16, 32 cycles per image; corresponding to 1.3, 2.6, 4, 5.3, 10.6 and 21.3 c/deg under our viewing conditions) with a bandwidth of one octave. The filtered noise was windowed by multiplying with a circular cosine of value one, falling to zero at the border over the space of 14 pixels, ensuring that letters did not distort beyond the borders of the padded image region. The amplitude of the filtered noise was then rescaled to have max / min values at 0.25, 0.5, 1, 1.5, 2, 2.5, 3, or 5 pixels; this controlled the strength of the distortion. One filtered noise sample controlled the horizontal pixel displacement, the other controlled vertical displacement (together giving the distortion map for the `griddata` algorithm).

Radial Frequency (RF) distortion: Here, the distortion map was created by modulating the distance of each pixel from the centre of the padded image sinusoidally (see Equation 1 in Dickinson et al., 2010, and 1E–G):

$$r'(\Theta) = r(1 + A\sin(\omega\Theta))$$

where r' is the distorted radius from the centre, r the undistorted radius, A is the amplitude of distortion (the proportion of the unmodulated distance from the centre), Θ is the polar angle and ω is the frequency of distortion (here 2, 3, 4, 5, 8 or 12 cycles in 2π radians). The phase of the modulation on each trial was drawn from a random uniform distribution spanning $[0, 2\pi]$. The amplitude of the distortion was set to one of 0.0075, 0.01, 0.0617, 0.1133, 0.1650, 0.2167, 0.2683 or 0.3200. The distortion map was windowed in a circular cosine as above, then the cosine and sine values were passed to `griddata` as the horizontal and vertical offsets.

To facilitate future modelling of our experiment, we pregenerated all images presented to observers (see below) and saved them to disk. In total we generated 1920 images: two distortion types (BPN, RF) × two conditions (flanked, unflanked; see below) × eight amplitudes × six frequencies, each repeated 10 times (i.e., 10 unique images were generated per cell). Target positions, letter identities and distortions were randomised on each repeat. In addition, we generated the same 1920 images *without* applying distortion to one of the target letters and saved them to disk. An image-based model of pattern recognition could be evaluated on the same stimuli as we have shown to our observers, using an undistorted "full-reference" image as a baseline (all images are provided online at http://dx.doi.org/10.5281/zenodo.48574).

2.1.4 Procedure

On each *unflanked* trial, observers saw the four target letters and indicated which letter was distorted. The letters subtended approximately 1.5×1.5 dva and were located above, below, right and left of fixation (see Figure 2A); letter identity at each location was randomly shuffled on each trial. The target letters were centred at a retinal eccentricity of 320 pixels (7.7 dva), and observers were instructed to maintain fixation on the central fixation cross (best for steady fixation from Thaler et al., 2013). The entire letter array was presented on a square background of maximum luminance (side length 1024 pixels or 24.3 dva); the remainder of the monitor area was set to mean grey. Letter strokes were set to minimum luminance (i.e. the letters were approximately 100% Michelson contrast). The letter array was presented for 150 ms (abrupt onset and offset), after which the screen was replaced with a fixation
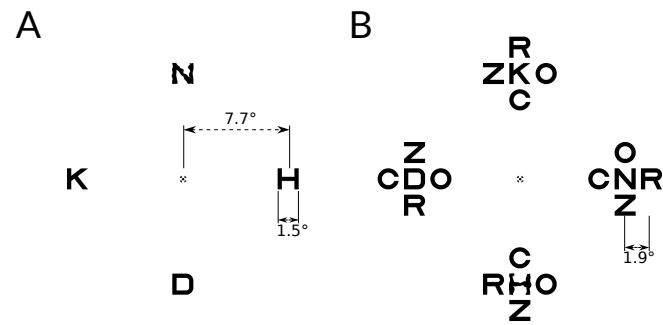
5

Figure 2: Example stimulus arrays. **A:** *An unflanked trial example. In this example the correct response is "above".* **B:** *A flanked trial example. The correct response is "below".*

cross on the same square bright background. The observer had up to 2000 ms to respond (a response triggered the next trial with ITI 100 ms), and received auditory feedback as to whether their response was correct.

On *flanked* trials (Figure 2B), four undistorted flanking letters the same size as the target were presented above, below, left and right of each target letter (centre-to-centre separation 1.9°, corresponding to approximately 0.25 of the eccentricity, well within the spacing of "Bouma's law"; Bouma (1970)). The arrangement of the four flanking letters was randomly determined on each trial.

Different distortion frequencies (six levels) and amplitudes (seven levels[1]) were randomly interleaved within a block of trials, whereas the distortion type (BPN or RF) and letter condition (unflanked or flanked) were presented in separate blocks. Each pairing of frequency and amplitude was repeated 10 times (corresponding to the unique images generated above), creating 420 trials per block. Breaks were enforced after every 70 trials. Blocks of trials were arranged into four-block sessions, in which observers completed one block of each pairing of distortion type and letter condition. Observers always started the session with an unflanked letter condition in order to familiarise them with the task[2]. Each session took approximately two hours. All observers participated in at least four sessions. Before the first block of the experiment observers completed 70 practice trials to familiarise themselves with the task. In total we collected 20,160 trials on each of the unflanked and flanked conditions.

2.1.5 Data analysis

Data from each experimental condition were fit with a cumulative Gaussian psychometric function using the *psignifit 4* toolbox for Matlab (Schütt et al., 2016), with the lower asymptote fixed to chance performance (0.25). The posterior mode of the threshold parameter (midpoint of the unscaled cumulative function) and 95% credible intervals were calculated using the default (weak) prior settings from the toolbox. The 95% credible intervals mean that the parameter value has a 95% probability of lying in the interval range, given the data and the prior. Psychometric function widths (slopes) either did not vary appreciably over experimental conditions (Experiment 1) or, when they did (Experiment 2), patterns of variation showed effects consistent with the threshold estimates. This paper therefore presents

---

[1] We generated stimuli for eight amplitudes but adjusted the sampling range after pilot testing to better sample the range of performance. All observers have done some trials at all amplitudes.

[2] Any practice effect should therefore improve performance in the flanked condition (this is not what we found).
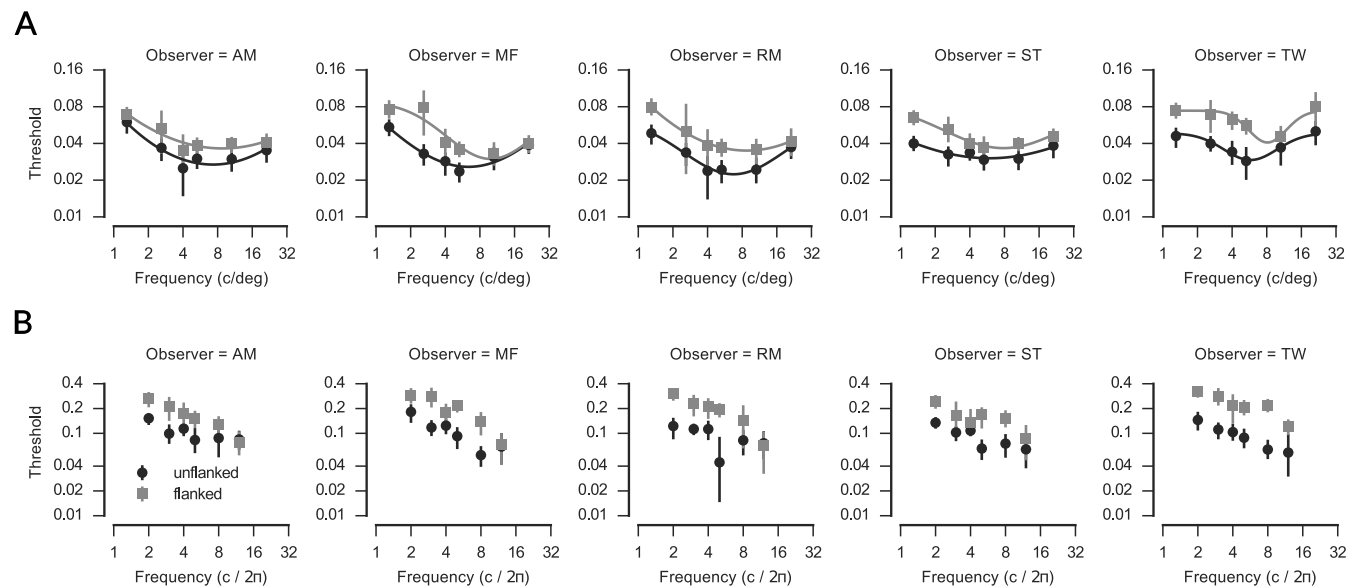
6

*Figure 3: Results of Experiment 1. **A:** Threshold amplitude for detecting letters distorted with BPN distortions, as a function of distortion frequency (c/deg) for five observers. Note both the x- and y-axes are logarithmic. Points show the posterior MAP estimate for the psychometric function threshold; error bars show 95% credible intervals. Thresholds are higher (observers are less sensitive to distortions) when flanking letters are present (light squares) compared to unflanked conditions (dark circles). Additionally, thresholds appear to show tuning, being lowest at approximately 6–8 c/deg. Curves show fits of a Gaussian function to the log frequencies and amplitudes (see text for details). **B:** Same as in A for RF distortion. Flanking letters again impair performance. Unlike in the BPN distortions, for RF distortions performance simply worsens for higher distortion frequencies. The reader can appreciate these results for themselves by examining how distortion visibility changes as a function of frequency in Figure 1D and G.*

only threshold data for brevity.

2.2 Results

Thresholds for detecting the distorted target letter are shown in Figure 3. For both distortion types, observers were less sensitive to letter distortion (thresholds were higher) when the target letters were surrounded by four flanking letters (grey squares) compared to when targets were isolated (black circles). This pattern is consistent with crowding.

Furthermore, we observe that the two distortion types (BPN and RF) show different dependencies on their respective frequency parameters (which are not themselves comparable). RF distortions become easier to detect the higher their frequency (c / 2π radians). BPN distortions show evidence of tuning, such that thresholds are lowest for frequencies in the range of 4–10 c/deg and rise for both lower and higher frequencies (note the log-log scaling in Figure 3). To quantify the shape of this tuning function we fit the log frequency and threshold data with a four-parameter inverted Gaussian (minimising the sum of squared errors with the BFGS method of Scipy's `minimize` function). Estimates of distortion frequency at which thresholds were lowest are shown for each observer in Figure 4. This procedure revealed a clear effect of flanking, such that when flanking letters were present, distortion sensitivity peaked at higher frequencies than when target letters were unflanked (flanked M = 8.69, SD = 0.80; unflanked M = 6.42, SD = 0.62; a difference
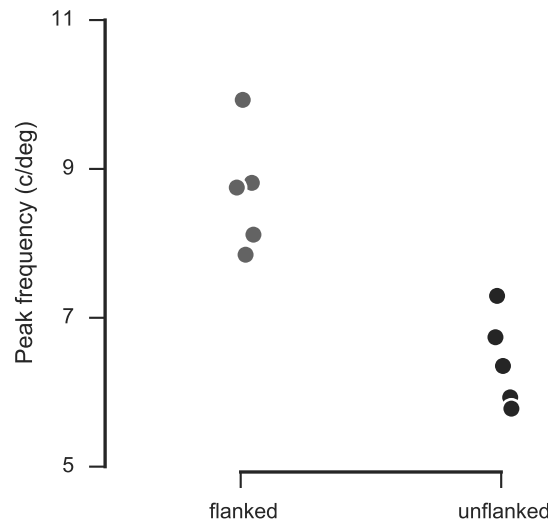
7

Figure 4: Peak frequency of lowest thresholds from the BPN distortion data (Figure 3), for each observer (points), in the flanked and unflanked conditions. Flanking letters cause sensitivity to shift to higher peaks. Points have been horizontally jittered to aid visibility.

of 0.44 octaves). A Bayesian paired t-test conducted using the free software JASP (version 7.1.12; Love et al., 2015; Rouder et al., 2012; Morey and Rouder, 2015) supported this conclusion, revealing a Bayes factor of 17.14 against the null model of no difference between the conditions (effect size median estimate 2.1, 95% credible interval 0.4 to 4.5, default prior settings). While the effect is therefore large compared to the relevant error variance, note that this computation ignores the precision with which the peak frequency is determined by the data, and so should be interpreted with some degree of caution.

The detectability of a given distortion will depend on the image content to which it is applied (for example, distorting a blank image region results in no image change). Performance indeed varied according to the target letter (Figure 5). On average across observers, it was easier to detect distortions applied to the letters K and H than the letters D and N, for both distortion types. Note however that the comparisons in (Figure 5) conflates both distortion sensitivity and response bias. Because each letter is presented on every trial (with the distortion applied to only one of the letters), an observer with a bias to choose a particular letter when in doubt would also serve to raise proportion correct performance (or thresholds). Thus, biases that are consistent across observers could also produce differences in letter performance. Measuring sensitivity to distortions in each letter while eliminating bias would require a forced-choice on individual letters (e.g. which of these "K"s is distorted?).

3   Experiment 2

Our first experiment showed that sensitivity to both BPN and RF distortions was reduced in the presence of undistorted flanking letters. Interestingly, our observers reported experiencing subjective "pop-out" in the flanked condition, such that the distorted letter appeared relatively more salient than the three undistorted targets by virtue of its contrast with neighbouring undistorted flankers. That is, the distorted letter strokes appeared subjectively more noticable when next to undistorted strokes. While the data quantitatively argue against such
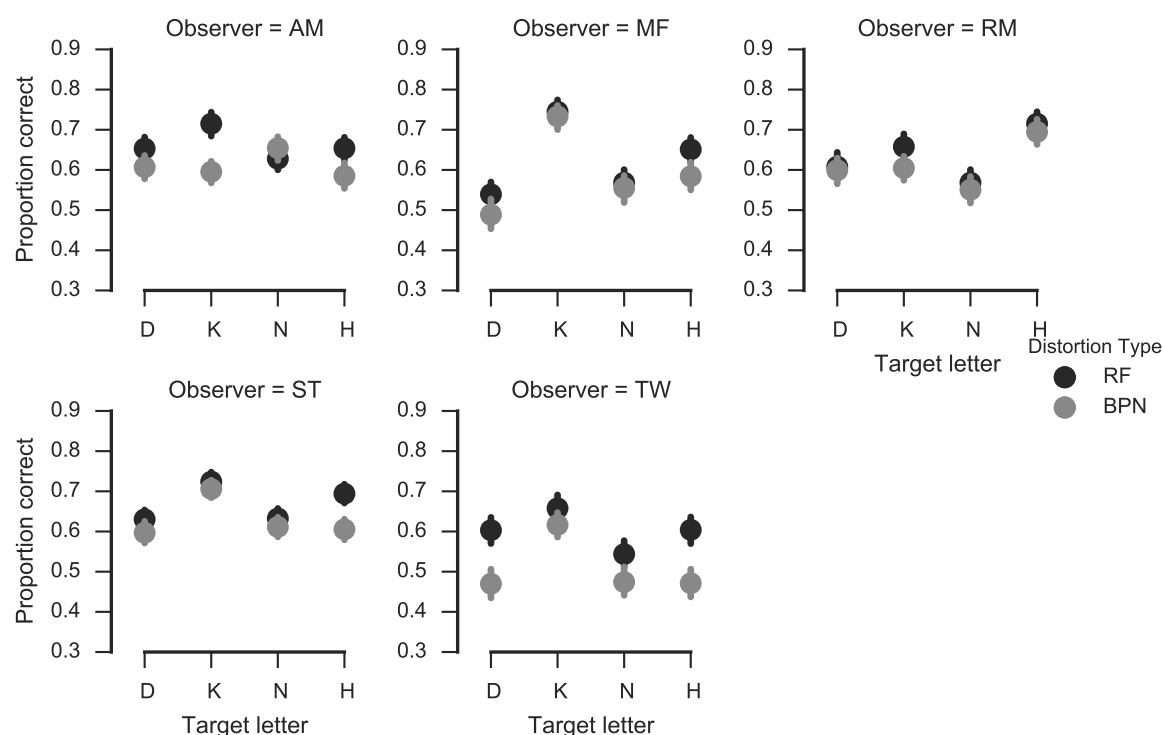
8

Figure 5: Average performance for each target letter and observer. Points show the proportion correct (error bars are bootstrapped 95% confidence intervals) for each target letter in each distortion type (colours), averaged over frequencies and amplitudes. The letters K and H generally show higher performance than D and N, for both distortion types. This could reflect either an interaction between letter shape and distortion (i.e. it is easier to discriminate distortions applied to the letter K), or biases in preferred letter.
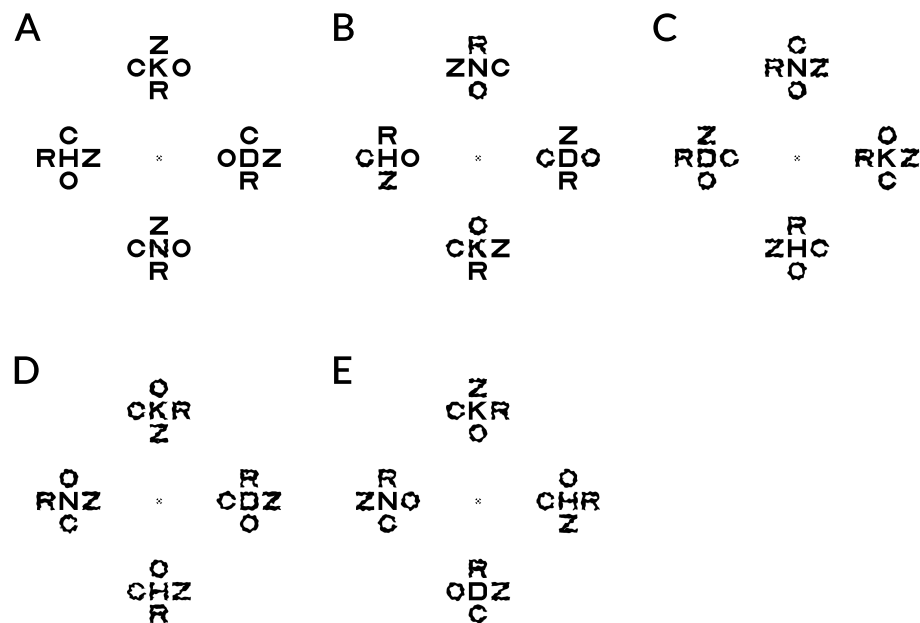
Figure 6: Example stimulus displays from Experiment 2 (all examples show the BPN distortion type at high distortion amplitudes). In Experiment 2a, observers detected the distorted middle letter when surrounded by zero (A), two (B) or four (C) distorted flankers. D: In Experiment 2b, observers indicated the undistorted middle letter surrounded by four distorted flankers. E: In Experiment 2c, flankers were always distorted at a highly-detectable distortion level. The correct response in panels A–E are down, left, down, left and right.

240 a pop-out effect (since flanking letters impaired performance), we nevertheless decided to
241 conduct a series of follow-up experiments to determine whether there was any dependence
242 of the thresholds on the kind of flankers employed. It is known from the crowding literature
243 that flankers more similar to the target cause stronger crowding (e.g. Bernard and Chung,
244 2011; Kooi et al., 1994); it is therefore plausible that distorted flankers would produce even
245 greater performance impairment.
246 We test this hypothesis in three related sub-experiments. Because we will directly com-
247 pare the data from each experiment, we present the similarities and differences in the ex-
248 perimental procedures first, followed by all data collectively. Three of the observers from
249 Experiment 1 (two authors plus one lab member) participated in these experiments; all
250 other experimental procedures were as in Experiment 1 except as noted below. As in Ex-
251 periment 1, all test images were pregenerated and saved along with undistorted reference
252 images to facilitate future modelling work.

3.1 Methods

3.1.1 Experiment 2a: varying the number of distorted flankers

255 This experiment was identical to Experiment 1, with the primary exception that in some
256 trials either two or four of the flanker letters in every letter array (above, left, below and
257 right) were also distorted (see Figure 6A–C). That is, observers reported which of the four
258 target letters was distorted, sometimes in the presence of distorted flankers. If distorted
259 targets pop out from undistorted flankers *and* undistorted targets pop out from distorted
260 flankers (*symmetrical popout*), we might expect that settings in which two of four flankers
261 are distorted would be hardest. In the case of no undistorted flankers (i.e. the same as the

10

flanked condition in Experiment 1), the distorted target pops out from the flankers. In the case of four distorted flankers, the *un*distorted targets pop out in three of the four possible locations, alerting the observer to the correct response by elimination. Finally, when two flanking letters are distorted, any differential pop-out signal is minimised because the non-target letter arrays contain two distorted letters whereas the letter array corresponding to the correct response contains three distorted letters. This account would therefore predict that thresholds in the two distorted flanker letter condition should be higher than those for zero or four distorted flankers.

In this experiment we selected one distortion frequency for each distortion type: 2.6 c/deg for the BPN and 4 c/2π for the RF distortions. Because our pilot testing indicated these tasks were more difficult than those in Experiment 1, we generated distortions at higher amplitudes than those in the first experiment: 0.024, 0.048, 0.072, 0.096, 0.120, 0.144, and 0.168 for BPN and 0.05, 0.125, 0.2, 0.275, 0.25, 0.425 and 0.5 for RF. Flanking letters were distorted with the same frequency and amplitude distortion as the target letter on every trial.

Trials of different distortion types (BPN, RF) and flanker conditions (zero, two or four distorted flankers) were presented in separate blocks in which each of the seven amplitudes were randomly interleaved. Ten unique images were created for each amplitude, each repeated three times to give 30 trials per amplitude (210 per block). Blocks of trials were arranged into six-block sessions, consisting of each distortion type and flanker condition in a random order for each observer. All observers participated two sessions, creating a total of 7560 trials.

3.1.2 Experiment 2b: detect the undistorted letter in the presence of distorted flankers

In Experiment 1, observers detected which of four letters was distorted when surrounded by four undistorted flanking letters. In Experiment 2b we examine the inverse task: to detect which middle letter is *un*distorted in the presence of four distorted flankers (Figure 6D). If distortion detection is symmetric, performance in this condition should be as good as in the zero distorted flanker condition of Experiment 2a. That is, distorted letters should pop out from undistorted flankers just as undistorted letters pop out from distorted flankers. The procedure was otherwise identical to Experiment 2a, with the exception that observers did two blocks (BPN and RF distortion types) of 210 trials (totalling 1260 trials).

3.1.3 Experiment 2c: flanker distortion at fixed high amplitude

In Experiments 2a and 2b, flanker distortions had the same amplitude as the target letter distortion. Therefore, for low target distortion amplitudes the flanker distortions may also be subthreshold. Popout, if it exists, may require detectable levels of distortion in the flanking elements. To test this question we repeated the four distorted flanker condition from Experiment 2a, with the exception that the flankers were distorted at a fixed amplitude that rendered distortions easily detectable (0.144 c/deg for BPN, 0.425 c/2π for RF; see Figure 6E). If popout requires suprathreshold distortions in flanking letters then sensitivity in this condition should be higher than the four distorted flanker condition from Experiment 2a (i.e. more similar to the zero distorted flanker condition for Experiment 2a). Observers performed at least two blocks, one for each distortion type (2520 trials total).
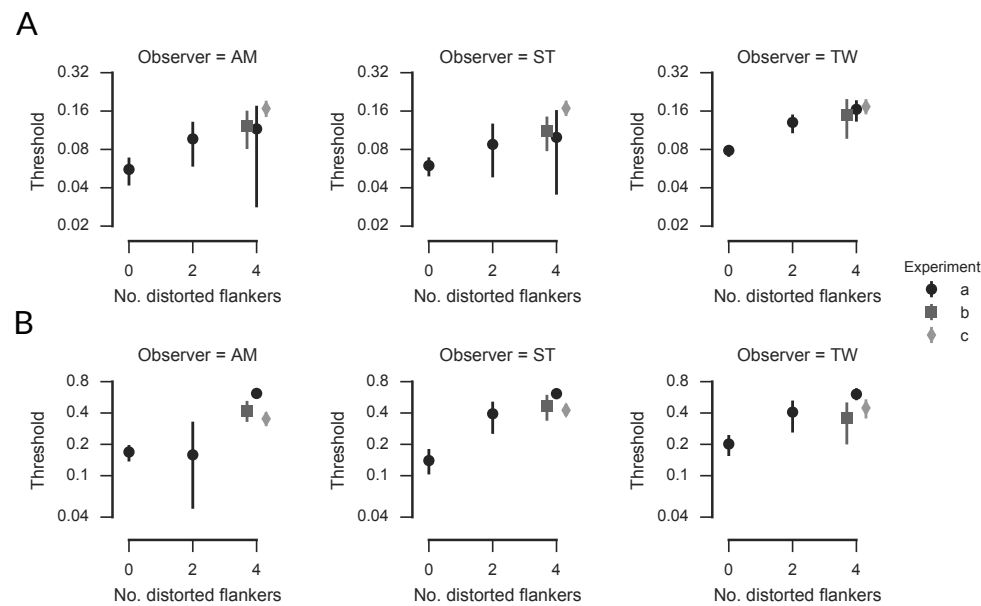
11

Figure 7: Results of Experiment 2. A: BPN distortions. Threshold amplitude for detecting target letter as a function of the number of distorted flankers, for three observers. Note the logarithmic y-axis. Points show the posterior MAP estimate for the psychometric function threshold; error bars show 95% credible intervals. Points for Experiments 2b and c have been shifted in the x direction to aid visibility; they both had four distorted flankers. B: Same as in A for RF distortion.

3.2   Results

Threshold levels of distortion are shown in Figure 7. The results for the BPN and RF distortions show qualitatively similar effects of the experimental conditions. First, thresholds increase as more flanking letters are distorted: detecting distortions in arrays with two or four distorted flankers is more difficult than when no flankers are distorted (Experiment 2a; Figure 7 circles). There is therefore no support for the prediction that thresholds would be higher in the two distorted flanker condition which, had it occurred, would be consistent with targets popping out from (un)distorted flankers in the zero and four distorted flanker conditions.

The results of Experiment 2b (Figure 7, squares) also provided no support for symmetrical popout. There was no evidence that detecting an undistorted target letter amongst four distorted flankers was as difficult as the zero distorted flanker condition of Experiment 2a; instead, detecting the undistorted target letter was similarly difficult as detecting a distorted target letter amongst four distorted flankers.

Finally, thresholds in Experiment 2c (Figure 7, diamonds) show that detecting a distorted letter amongst four distorted flankers requires substantially more distortion amplitude than those with no distorted flankers, despite the flanker distortions being always easily detectable.

4   Discussion

We have measured human sensitivity to geometric distortions of letter stimuli presented to the peripheral retina. For two types of distortion, Experiment 1 showed that distortion sensitivity is reduced when target letters are surrounded by task-irrelevant flankers. This

result is consistent with crowding (Bouma, 1970). Crowding has previously been shown to impair both letter identification and orientation sensitivity. Our results could be considered to probe an intermediate level of representation: geometric distortions can change the contours of these simple but highly familiar shapes. Detecting deviations from expected shape must involve local orientation processing, and strong distortions of letter shape impair letter identification (Wiecek et al., 2014). It is therefore unsurprising that the presence of surrounding flanking letters impairs geometric distortion detection; our results demonstrate this impairment and chart its strength in two distortion types.

We measured sensitivity to two distortion types in order to provide two physically distinct image changes to spur future modelling efforts. The frequency and amplitude parameters for each distortion type represent different physical image changes. Radial frequency distortions are highly correlated both tangentially and radially, whereas BPN distortions are not, and these correlations will interact with the original structure of the letter. Furthermore, BPN distortions of sufficient amplitude (when the pixel shift exceeds half the distortion wavelength) will cause reversals in pixel positions, producing "speckling" at high frequencies but leaving the mean position of low frequency components unchanged (see for example Figure 1D, the highest amplitude distortions for the two highest frequencies). Each distortion type produces different patterns of human sensitivity as a function of its distortion parameters, and a direct comparison between them was not the goal of this paper.

For BPN distortions, Experiment 1 revealed that distortion sensitivity is tuned to mid-range distortion frequencies (approximately 6–9 c/deg). This tuning is likely to reflect sensitivity to the speckling mentioned above: detecting high frequency distortions requires detecting high frequency speckles, which are difficult to see in the periphery due to acuity loss [3]. Thresholds therefore rise again compared to mid-frequency distortions, which observers can detect well before speckling occurs. Experiment 1 also showed that when flankers are present, peak sensitivity shifts to higher frequencies than when flankers are absent. This could be caused because flanking letters selectively reduce sensitivity to position changes at lower spatial scales, or because flanking letters increase sensitivity to higher-frequency speckles. Given that there is no plausible mechanism that might support the latter possibility, we favour the former.

How do the BPN distortion peaks in our data fit with previous studies employing BPN distortions? In Wiecek et al. (2014), letters of different sizes were presented foveally, and participants identified the letter after BPN distortion. Letter identification performance showed different tuning for distortion frequency at different letter sizes. Filtering with a peak frequency of 8 c/deg produced poorest identification performance for letters subtending 0.33 deg. These results fit with our data, if we assume that when a distortion is maximally detectable (our experiment) it maximally reduces letter identification (Wiecek et al. (2014)); the difference in letter size likely reflects a size scaling constant in detectability as letters move away from the fovea (Chung et al., 2002; Song et al., 2014). Observers in Bex (2010) detected BPN distortions introduced into one quadrant of natural scenes. He found that observers were maximally sensitive to distortions of approximately 5 c/deg, and that these peaks were relatively stable for distortions centred at retinal eccentricities of 1.5, 2.8 and 5.6 deg. These estimates appear to be at the lower bound of those we observe here, suggesting that distortion detection sensitivity in letter stimuli peaks at higher spatial scales than detecting distortions of natural scene content (though note that the results of Wiecek et al. (2014) imply that the peaks we observe will also depend on letter size).

Our Experiment 1 showed that flanking letters reduced sensitivity to letter distortions.

---

[3]We credit Peter Bex for pointing out the likely relevence of speckling to the observed tuning.

13

In the follow-up studies of Experiment 2 we found that this impairment became more severe when flanking letters were themselves distorted. An averaging account of this task might predict that undistorted contours of the target letters will appear more distorted by virtue of lying nearby the distorted flankers (e.g. in Experiment 2c), thus reducing the perceived difference between distorted and undistorted target letters and reducing performance. In other words, crowding makes the targets *appear* more like the flankers (Greenwood et al., 2010). Increasing the number of distorted flankers (as in Experiment 2a) also reduces the difference signal available, if averaging depends on the number of nearby distorted contours. In contrast, a substitution model holds that observers encode the target and flanker features (in this case, distortedness) accurately, but will sometimes confuse the flankers and the target and thus report the flanker characteristic on some trials. Our results therefore seem more consistent with averaging than with substitution models of crowding. A substitution model could account for our results if the substituting observer *only* has access to the distortedness of the four letters perceived to be at the target location (whether or not these are targets or substituted flankers). A more realistic assumption would be that the observer can tell whether flanker letters are also distorted. In this case, the observer should be able to respond correctly because they would know which array of five letters contains the oddball letter, irrespective of whether a flanker is substituted for a target. While our data therefore seem to favour an averaging model, we would like to stress that the present study was not designed to discriminate between these accounts of crowding. Recent work shows that both averaging- and substitution-like errors can arise from a simple population coding model and decision criterion, at least for the case of orientation discrimination (Harrison and Bex, 2015).

5   Conclusion

Taken together, the pattern of results presented here provide a challenge for models of 2D form processing in humans. A successful model of form discrimination would need to explain sensitivity to two distinct distortion types, the dependence of distortion sensitivity on flanking letters, and the dependence on the type of flanking letters (distorted flankers reduce sensitivity). Directly comparing the BPN and RF distortions would require an image-based similarity metric that captured the perceptual size of the distortions on a common scale. One test of such a similarity metric would be to rescale the results of the BPN and RF data reported here such that the different sensitivity patterns as a function of distortion frequency overlap. We have provided our raw data and images of the stimuli used in these experiments (http://dx.doi.org/10.5281/zenodo.48574) to facilitate future efforts along these lines.

6   Acknowledgements

14

References

Andriessen, J. J. and Bouma, H. (1975). Eccentric vision: Adverse interactions between line segments. *Vision Research*, 16(1):71–78.

Bernard, J. B. and Chung, S. T. L. (2011). The dependence of crowding on flanker complexity and target-flanker similarity. *Journal of Vision*, 11(8):1.

Bex, P. J. (2010). (In) sensitivity to spatial distortion in natural scenes. *Journal of Vision*, 10(2):23.1–15.

Bouma, H. (1970). Interaction effects in parafoveal letter recognition. *Nature*, 226(5241):177–178.

Brainard, D. H. (1997). The Psychophysics Toolbox. *Spatial Vision*, 10(4):433–436.

Chung, S. T. L., Legge, G. E., and Tjan, B. S. (2002). Spatial-frequency characteristics of letter identification in central and peripheral vision. *Vision Research*, 42(18):2137–2152.

Dickinson, J. E., Almeida, R. A., Bell, J., and Badcock, D. R. (2010). Global shape aftereffects have a local substrate: A tilt aftereffect field. *Journal of Vision*, 10(13):5.

Greenwood, J. A., Bex, P. J., and Dakin, S. C. (2009). Positional averaging explains crowding with letter-like stimuli. *Proceedings of the National Academy of Sciences of the United States of America*, 106(31):13130–13135.

Greenwood, J. A., Bex, P. J., and Dakin, S. C. (2010). Crowding changes appearance. *Current Biology*, 20(6):496–501.

Harrison, W. J. and Bex, P. J. (2015). A Unifying Model of Orientation Crowding in Peripheral Vision. *Current Biology*, 25(24):3213–3219.

Herzog, M. H., Sayim, B., Chicherov, V., and Manassi, M. (2015). Crowding, grouping, and object recognition: A matter of appearance. *Journal of Vision*, 15(6).

Hole, G. J., George, P. A., Eaves, K., and Rasek, A. (2002). Effects of geometric distortions on face-recognition performance. *Perception*, 31(10):1221–1240.

Kingdom, F. A., Field, D. J., and Olmos, A. (2007). Does spatial invariance result from insensitivity to change? *Journal of Vision*, 7(14):11.

Kleiner, M., Brainard, D. H., and Pelli, D. G. (2007). What's new in Psychtoolbox-3? *Perception*, 36(ECVP Abstract Supplement).

Kooi, F. L., Toet, A., Tripathy, S. P., and Levi, D. M. (1994). The effect of similarity and duration on spatial interaction in peripheral vision. *Spatial Vision*, 8(2):255–279.

Levi, D. D. M., Klein, S. A., and Aitsebaomo, A. P. (1985). Vernier acuity, crowding and cortical magnification. *Vision Research*, 25(7):963–977.

Love, J., Selker, R., Marsman, M., Jamil, T., Dropmann, D., Verhagen, A. J., Ly, A., Gronau, Q. F., Smira, M., Epskamp, S., Matzke, D., Wild, A., Knight, P., Rouder, J. N., Morey, R. D., and Wagenmakers, E.-J. (2015). JASP.

453   Morey, R. D. and Rouder, J. N. (2015). BayesFactor.

454   Parkes, L., Lund, J., Angelucci, A., Solomon, J. A., and Morgan, M. J. (2001). Compulsory
455   averaging of crowded orientation signals in human vision. *Nature Neuroscience*, 4(7):739–
456   744.

457   Pelli, D. G. (1997). The VideoToolbox software for visual psychophysics: transforming num-
458   bers into movies. *Spatial Vision*, 10(4):437–442.

459   Rouder, J. N., Morey, R. D., Speckman, P. L., and Province, J. M. (2012). Default Bayes
460   factors for ANOVA designs. *Journal of Mathematical Psychology*, 56(5):356–374.

461   Rovamo, J., Mäkelä, P., Näsänen, R., and Whitaker, D. (1997). Detection of geometric im-
462   age distortions at various eccentricities. *Investigative Ophthalmology & Visual Science*,
463   38(5):1029–1039.

464   Schütt, H., Harmeling, S., Macke, J. H., and Wichmann, F. A. (2016). Painfree and accu-
465   rate Bayesian estimation of psychometric functions for (potentially) overdispersed data.
466   *Vision Research*.

467   Sloan, L. L. (1959). New test charts for the measurement of visual acuity at far and near
468   distances. *American Journal of Ophthalmology*, 48:807–813.

469   Song, S., Levi, D. M., and Pelli, D. G. (2014). A double dissociation of the acuity and crowd-
470   ing limits to letter identification, and the promise of improved visual screening. *Journal*
471   *of Vision*, 14(5):3, 1–37.

472   Spence, M. L., Storrs, K. R., and Arnold, D. H. (2014). Why the long face? The importance
473   of vertical image structure for biological "barcodes" underlying face recognition. *Journal*
474   *of Vision*, 14(8):25–25.

475   Stojanoski, B. and Cusack, R. (2014). Time to wave good-bye to phase scrambling: Creat-
476   ing controlled scrambled images using diffeomorphic transformations. *Journal of Vision*,
477   14(12):6.

478   Strasburger, H. (2014). Dancing letters and ticks that buzz around aimlessly: On the origin
479   of crowding. *Perception*, 43(9):963–76.

480   Thaler, L., Schütz, A. C., Goodale, M. A., and Gegenfurtner, K. R. (2013). What is the best
481   fixation target? The effect of target shape on stability of fixational eye movements. *Vision*
482   *Research*, 76:31–42.

483   Toet, A. and Levi, D. M. (1992). The two-dimensional shape of spatial interaction zones in
484   the parafovea. *Vision Research*, 32(7):1349–1357.

485   Wiecek, E., Dakin, S. C., and Bex, P. (2014). Metamorphopsia and letter recognition. *Journal*
486   *of Vision*, 14(14):1–1.

487   Wilkinson, F., Wilson, H. R., and Habak, C. (1998). Detection and recognition of radial
488   frequency patterns. *Vision Research*, 38(22):3555–3568.